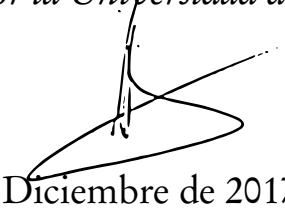
Evaluación de la magnitud y la calidad de la evidencia de los estudios secundarios relacionados con la psoriasis: utilidad de las revisiones sistemáticas, los meta-análisis en red y los estudios meta-epidemiológicos

Francisco José Gómez García

Departamento de Medicina, Dermatología y Otorrinolaringología

Trabajo presentado para optar al grado de
Doctor por la Universidad de Córdoba



Diciembre de 2017

TITULO: *Evaluación de la magnitud y la calidad de la evidencia de los estudios secundarios relacionados con la psoriasis: utilidad de las revisiones sistemáticas, los meta-análisis en red y los estudios meta-epidemiológicos*

AUTOR: *Francisco José Gómez García*

Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

A mi familia.

DON ANTONIO VÉLEZ GARCÍA NIETO, PROFESOR TITULAR DEL DEPARTAMENTO DE MEDICINA, DERMATOLOGÍA y OTORRINOLARINGOLOGÍA DE LA UNIVERSIDAD DE CÓRDOBA Y DIRECTOR DE LA UGC DE DERMATOLOGÍA DEL HOSPITAL UNIVERSITARIO REINA SOFÍA,

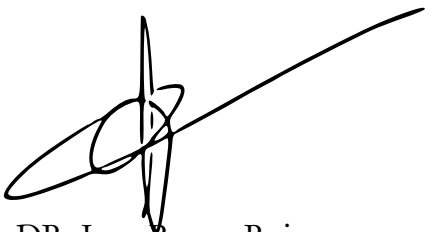
DON JUAN RUANO RUIZ, DOCTOR EN MEDICINA POR LA UNIVERSIDAD DE CÓRDOBA Y RESPONSABLE DEL GRUPO DE INVESTIGACIÓN GE03 'ENFERMEDADES INFLAMATORIAS CUTÁNEAS INMUNOMEDIADAS' DEL INSTITUTO MAIMÓNIDES DE INVESTIGACIÓN BIOMÉDICA DE CÓRDOBA

Hacemos constar:

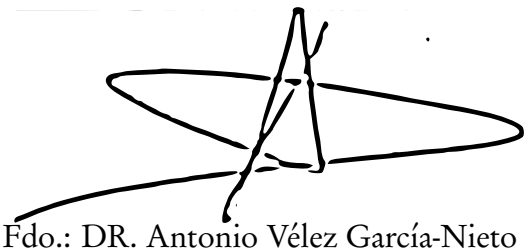
Que el trabajo titulado: ***"Evaluación de la magnitud y la calidad de la evidencia de los estudios secundarios relacionados con la psoriasis: utilidad de las revisiones sistemáticas, los meta-análisis en red y los estudios meta-epidemiológicos"*** ha sido realizado por **Don Francisco José Gómez García**, Licenciado en Medicina y Cirugía, bajo nuestra dirección, dentro del programa de Doctorado "Biomedicina" y desarrollado conjuntamente en la UGC de Dermatología del Hospital Universitario Reina Sofía y en el Instituto Maimónides de Investigación Biomédica de Córdoba (IMIBIC).

A nuestro juicio reúne los méritos suficientes para ser defendido ante el tribunal correspondiente y poder optar al grado de Doctor.

Córdoba, 4 de diciembre de 2017



Fdo.: DR. Juan Ruano Ruiz



Fdo.: DR. Antonio Vélez García-Nieto

TÍTULO DE LA TESIS: "Evaluación de la magnitud y la calidad de la evidencia de los estudios secundarios relacionados con la psoriasis:utilidad de las revisiones sistemáticas, los meta-análisis en red y los estudios meta-epidemiológicos."

DOCTORANDO/A: Francisco José Gómez García

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

D. Francisco José Gómez García presenta un trabajo original cuyo aspecto nuclear es la síntesis de la evidencia. El abordaje se ha llevado a cabo desde un doble punto de vista. En primer lugar, mediante la realización de una revisión sistemática y meta-análisis en red sobre la eficacia y seguridad a corto plazo de los tratamientos biológicos autorizados en el tratamiento de la psoriasis en placas moderada-severa. En segundo lugar, a través de la evaluación científica de las revisiones sistemáticas publicadas sobre psoriasis. Para ello, el doctorando ha adquirido los conocimientos y habilidades sobre la metodología relacionada con la conducción y la notificación de las revisiones sistemáticas, así como para el manejo de las herramientas de evaluación de la calidad científica de este tipo de documentos.

Los resultados obtenidos en este trabajo han sido publicados en tres artículos de dos revistas científicas de *reconocido prestigio internacional en este campo de la investigación tanto dermatológica* [*British Journal of dermatology* (D1/Q1:5/63)] como de las ciencias y servicios de atención médica [*Journal of Clinical Epidemiology* (D1/Q1:6/90)]. Serán defendidos mediante el sistema de Tesis por compendio de artículos, ya que constituyen una misma unidad temática de objetivos y resultados progresivos. Además, han dado lugar a otros trabajos científicos en los que el doctorando es autor o coautor, todos ellos publicados en revistas situadas en Q1 y cuyas temáticas se integran en el área de estudio de la síntesis del conocimiento.

La tesis doctoral presentada se enmarca dentro del proyecto de investigación del Instituto Carlos III "***Ensayo clínico aleatorizado y multicéntrico para evaluar, en términos de coste-eficiencia, un modelo predictivo de respuesta a fármacos anti-TNFs (ICI400136)"*** en psoriasis moderada severa, cuyo IP es el Codirector de la Tesis Dr Juan Ruiz Ruano. Además sus resultados apoyan el desarrollo de un nuevo proyecto que ha obtenido financiación pública competitiva "***Proyecto Éaco: Desarrollo y Validación de Un Instrumento Tecnológico en Red para el Apoyo en la Toma de Decisiones en la Práctica Clínica (PIN-0316-2017)"***

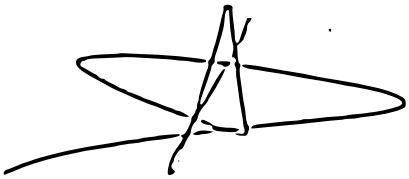
Finalmente, cabe destacar la formación técnica y científica alcanzada por el doctorando que ha sido excelente. El desarrollo de la tesis le ha permitido adquirir conocimientos teóricos y metodológicos que lo capacitan para desarrollar nuevas hipótesis y participar activamente en la redacción y coordinación de nuevos artículos científicos y proyectos de investigación.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 4 de diciembre de 2017

Firma del/de los director/es

Fdo.: Dr. Juan Ruano Ruiz



Fdo.: Dr. Antonio Vélez García-Nieto

Agradecimientos

Me gustaría agradecer a todos los que en algún momento se tomaron la molestia de enseñarme algo, una parte de ellos esta recogida en estas páginas.

A mis directores de tesis, al Dr. Ruano por hacerme sentir la ciencia con el fervor que yo la imaginaba cuando era un niño, por guiarme a veces como un padre, por mostrar esa responsabilidad e ímpetu en el trabajo que te empuja a escalar la mayor pendiente. Al Dr.Velez por haber confiado en mí, haberme dado la oportunidad de desarrollar este trabajo a su lado y estar siempre dispuesto con esa sensación de inmediatez que tanto tranquiliza.

A Macarena, Jesus, Juanlu, Pedro, Beatriz, las Cármenes y todas las personas que hacen del GE03 un equipo donde el interés de uno es el de todos. Este trabajo es nuestro.

A los que me enseñaron sobre dermatología, especialmente al Dr Moreno, a los que aprendí con ellos y a todos los compañeros con los que he tenido la oportunidad de trabajar. Todos, de alguna forma, han estimulado ideas de esta tesis.

A mis amigos, a Rafa que tanto empeño pone en lo que le digo. A Chacón, Lola, Paco, Sonia, Fernando, porque el tiempo que me relajo con ellos es una parte imprescindible para afrontar luego el trabajo.

A Trini y José porque vuestra disposición y cariño me han acompañado todo el tiempo dedicado a este trabajo. Muchas de las líneas os pertenecen.

A mis abuelos, porque me enseñasteis cosas complicadas de la vida de la manera mas simple y yo sin darme cuenta las aprendía siendo feliz a vuestro lado. Cuánto os echo de menos, a veces sin saberlo.

A Lola quien me ha dado y de quien espero lo mejor, porque tu ejemplo me enseñó y me sirve, y me haces sentir que esto es también tuyo y porque aún nos queda tanto que

pienso que solo vamos por el principio.

A Fran, porque la vida es un sueño sobre todo cuando me llamas para dormirlo contigo, porque las mejores ideas de este trabajo las has dibujado con tu dedo en mi cara y tienen el eco de tu risa y de tu llanto. Porque haces que mi mayor reto sea mi mejor placer.

A mis padres, que me lo han dado todo y con los que siento que mi única deuda es poder provocar en mi hijo la emoción de gratitud que me invade cuando pienso en vosotros. Con ella nunca estaré solo. Este trabajo es sobre todo vuestro.

Resumen

Introducción. La psoriasis es una enfermedad inflamatoria crónica de alta prevalencia cuyas formas moderadas y severas se asocian a una reducción de la calidad de vida y mayor comorbilidad. En relación a los tratamientos clásicos, los fármacos biológicos muestran elevada eficacia y mejor perfil de seguridad. Sin embargo, no están exentos de riesgo y sus elevados costes amenazan el sostenimiento del tratamiento de estos pacientes dentro de los sistemas públicos de salud. En este contexto, es especialmente importante que las decisiones de los clínicos y los gestores sanitarios se basen en las mejores pruebas científicas. Las revisiones sistemáticas, acompañadas en ocasiones de un meta-análisis, constituyen los documentos estándar para sintetizar la evidencia. El desarrollo de recomendaciones metodológicas y de herramientas para la conducción, notificación y el control de su calidad ha permitido minimizar el grado de incertidumbre de las estimaciones que establecen sobre los efectos analizados, mejorando así la eficiencia de las recomendaciones derivadas. Sin embargo, se ha observado en los últimos años un crecimiento exponencial de revisiones sistemáticas conflictivas. Conocer la calidad metodológica y el riesgo de sesgo de estas publicaciones en psoriasis permitirá seleccionar sólo las de mayor calidad científica.

Objetivos. En el presente proyecto de investigación nos propusimos una serie de objetivos primarios y secundarios que abordaríamos mediante dos estrategias. En primer lugar, llevamos a cabo una **revisión sistemática** y un **meta-análisis en red** sobre el tratamiento a corto plazo de la psoriasis en placas moderada-severa en adultos con fármacos biológicos. En segundo lugar, **evaluamos la calidad metodológica** y el **riesgo de sesgo** de las revisiones sistemáticas y los meta-análisis publicados sobre psoriasis, incorporando datos y metadatos relacionados con las revisiones, las revistas y los autores, con el fin de elaborar modelos predictivos que pueden servir de ayuda en la toma de decisiones.

Metodología. Para la conducción y notificación de las revisiones sistemáticas se empleó el manual de la Colaboración Cochrane y las recomendaciones **PRISMA** (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), respectivamente . La evaluación de la calidad metodógica y del riesgo de sesgo de las revisiones sistemáticas se realizó con **AMSTAR** (*A Measurement Tool to Assess the Methodological Quality of Systematic Reviews*) y **ROBIS** (*Risk Of Bias In Systematic Reviews*), respectivamente. En la síntesis se aplicó el

metaanálisis en red para la comparación de los fármacos y el análisis de regresión logística, análisis de componente principal y análisis de clustering en los estudios metaepidemiológicos.

Resultados. Los resultados derivados de los análisis anteriores han sido publicados como tres artículos originales en revistas de alto impacto científico (todas en el primer decil de su área temática). En resumen, tras publicar un protocolo *a priori* en PROSPERO, se llevó a cabo la revisión sistemática y el metaanálisis en red, incluyendo 27 ensayos clínicos aleatorizados y controlados. Infliximab 5 mg/Kg cada 8 semanas y secukinumab 300 mg cada 4 semanas fueron los tratamientos más eficaces para los resultados de PASI 75 y PASI 90 en la semana 10-16, respectivamente. Además, ambos tratamientos presentaron el mayor riesgo de tener al menos un efecto adverso (42.1 %) o al menos una infección (52.2%), respectivamente. No se encontraron diferencias en los acontecimientos adversos severos. Ustekinumab 90 mg cada 12 semanas fue el fármaco con mejor perfil riesgo beneficio para las medidas estudiadas. La calidad de la evidencia para los valores de eficacia fue alta en el caso de ustekinumab y modera o baja para el resto. Para los datos de seguridad resultó ser, en general, baja o muy baja. Respecto a los estudios metaepidemiológicos relacionados con la evaluación y predicción de la calidad metodológica, se incluyeron 220 revisiones sistemáticas publicadas por 741 autores de 520 instituciones diferentes procedentes de 32 países. Tan sólo el 17% de las revisiones presentaron alta calidad metodológica según AMSTAR. La inclusión de metaanálisis (OR 6.22, IC95% 2.78-14.86), la financiación por instituciones acadḿicas (OR 2.90, IC95% 1.11-7.89), un número elevado de autores con conflicto de intereses (OR 0.90, IC95% 0.82-0.99), el factor de impacto de la revista (OR 2.14, IC95% 1.05-6.67), el factor de impacto a 5 años (OR 1.34, IC95% 1.02-1.40) y el número de páginas (OR 1.08, IC95% 1.02-1.15) fueron predictoras de alta calidad metodológica. En referencia al análisis efectuado para comparar AMSTAR y ROBIS, se evaluaron 139 revisiones sistemáticas sobre intervenciones en psoriasis, de las que únicamente el 22.3% presentaron alta calidad metodológica y un 14% bajo riesgo de sesgo. El porcentaje de revisiones sistemáticas de alta calidad con alto riesgo de sesgo fue mayor del 50%. Los componentes que mejor explicaron la variabilidad de los resultados de AMSTAR y ROBIS están relacionados con la evaluación de los estudios primarios de las revisiones sistemáticas.

Abstract

Background. Psoriasis is a highly prevalent chronic inflammatory skin disease whose moderate and severe forms are associated with a reduction of quality of life and a greater comorbidity. In relation to the classic treatments, biologics show high efficacy and better safety profile. However, they are not risk-free and their high costs threat to sustain the treatment of these patients within the public health systems. In this context, it is especially important that decisions taken by clinicians and health managers will be supported on the best evidence. Systematic reviews and a meta-analysis constitute the standard documents for summarizing scientific evidence. Methodological instruments and protocols used in the development, notification, and quality control of these documents have allowed to reduce the degree of uncertainty of drug effect estimations, thus improving the efficiency of the derived recommendations. However, it has been observed in recent years an exponential growth of systematic reviews of dubious scientific quality. By assessing the methodological quality and the risk of bias of these publications, we can select only those of the highest scientific quality.

Objectives. In the present research project we proposed a series of primary and secondary objectives that were achieved through two different strategies. First, we carried out a systematic review and a network meta-analysis on the short-term effect of biologics for the treatment of moderate-severe plaque psoriasis in adults. Second, we assessed the methodological quality and the risk of bias of systematic reviews and meta-analyses published on psoriasis, by incorporating data and metadata related to articles, journals and authors, in order to develop a predictive model that can be helpful in decision-making processes.

Methodology. We followed the Cochrane Handbook for Systematic Reviews on interventions and the **PRISMA** (*Preferred reporting items for Systematic reviews and Meta-Analyses*) recommendations to for systematic reviews conduction and notification respectively. The evaluation of the methodological quality and the risk of bias of the systematic reviews were performed with **AMSTAR** (*A Measurement Tool to Assess the Methodological quality of Systematic reviews*) and **ROBIS** (*Risk of Bias In (Systematic reviews*) tools, respectively. For metaepidemiological analyses different analytic approach were followed: multinomial logistic regresion models, principal component analysis, clustering analysis.

Results. The results derived have been included in three original articles published in high-impact scientific journals (all ranked in the 10th percentile). In summary, after publishing *a priori* protocol in PROSPERO, it was carried out a systematic review and network meta-analysis, including 27 randomized-controlled clinical trials. Infliximab 5

mg/Kg every 8 weeks and secukinumab 300 mg every 4 weeks were the most effective treatments in week 10-16 accounting for PASI 75 and PASI 90 results respectively. In addition, both treatments presented the highest risk of having at least one adverse effect (42.1 %) or at least one infection (52.2 %), respectively. No differences were found in severe adverse events among the treatments. Ustekinumab 90 mg Every 12 weeks was the treatment with the best risk vs benefit profile. The quality of the evidence for efficacy estimates was high in the case of ustekinumab and moderate or low for the rest drugs. For the safety, quality of the evidence was, in general, low or very low. With respect to the metaepidemiological studies related to the evaluation and prediction of the methodological quality, 220 systematic reviews published by 741 authors of 520 different institutions from 32 countries were included. Only 17% of the reviews presented high methodological quality according to AMSTAR tool. The inclusion of a metan alisys (OR 6.22, (IC95% 2.78-14.86), funding by academid institutions (OR 2.90, IC95% 1.11-9.19), low number of authors with conflict of interests (OR 0.90, IC95% 0.82-0.99), a high journal impact factor (OR 2.14, IC95% 1.05-6.67), a high 5 years journal impact factor (OR 1.34, IC95% 1.02-1.40), and a high number of pages (OR 1.08, IC95% 1.02-1.15) were predictors of high methodological quality. In reference to the analysis carried out to compare AMSTAR and ROBIS, we evaluated 139 systematic reviews on interventions in psoriasis, of which only 22.3% presented a high methodological quality and 14 % a low risk of bias. The percentage of systematic reviews with high methodological quality and with high risk of bias was greater than 50%. The components that best explained the variability of the results of AMSTAR and ROBIS are related with the evaluation of the primary studies of systematic reviews.

Indice

Introducción

- **I. Psoriasis y medicina basada en la evidencia.**
- **II. Documentos de síntesis de la evidencia.**
  - II.a. Del arte de curar a la medicina basada en la evidencia.
  - II.b. Elaboración de revisiones sistemáticas.
  - II.c. Síntesis de los resultados de revisiones sistemáticas.
  - II.d. Notificación de los resultados de revisiones sistemáticas y meta-análisis.
- **III. Metaepidemiología y control de calidad de los documentos de síntesis de la evidencia.**
- **IV. Guías de práctica clínica: recomendaciones basadas en la evidencia para la toma de decisiones.**

I. Psoriasis y medicina basada en la evidencia.

La psoriasis es una enfermedad inflamatoria crónica, genéticamente compleja e inmunológicamente mediada. Su prevalencia alcanza del 1,5 al 3% de la población [1]. La forma clínica más frecuente es la psoriasis en placas cuyas lesiones se caracterizan por eritema, descamación y grosor. La gravedad de las lesiones individuales asociada con la extensión de la enfermedad definen la severidad de la psoriasis. Las formas moderadas-severas representan el 25% [2], asocian comorbilidades entre las que destaca la artritis psoriásica en hasta un 40% y presentan una mayor prevalencia de factores de riesgo cardiovascular (40-50%), lo que se asocia con el aumento de eventos cardiovasculares mayores (infarto de miocardio, accidente cerebrovascular o muerte por estas causas) [3]. Con respecto a la calidad de vida, un 75% de estos pacientes ven limitado su funcionamiento mental o físico en sus actividades diarias y hasta un 20% han contemplado el suicidio [4]. Todo ello justifica que el impacto sea similar al producido por el cáncer o enfermedades inflamatorias crónicas como la artritis.

Clásicamente se han empleado como tratamientos sistémicos la fototerapia, el metotrexato, la ciclosporina o el acitretino, pero la toxicidad acumulada de estos fármacos limita su uso. Posteriormente, un mejor conocimiento de la inmunopatogenia de la enfermedad ha permitido el desarrollo de nuevos fármacos. Estas terapias, denominadas biológicas, se sintetizan a partir de productos de organismos vivos. Tienen una acción muy selectiva mediante el bloqueo específico de citoquinas como el TNF-alfa (etanercept, infliximab, adalimumab) , la subunidad p40 de la interleucina (IL)-12 e IL-23 (ustekinumab) o la IL-17 (secukinumab, ixekizumab), entre otras. Presentan una eficacia elevada y permiten, a priori, el tratamiento a más largo plazo. Pero muestran un perfil de efectos secundarios potencialmente graves y unos costes elevados. En este sentido, se ha medido un aumento del 30% del gasto en el tratamiento de la psoriasis desde la aparición de los fármacos biológicos [5], lo que unido a las costes indirectos de la enfermedad como disminución del rendimiento, absentismo laboral o los derivados del

Fig. 1 La psoriasis en placas es la forma más frecuente de psoriasis.

tratamiento de acontecimientos adversos y comorbilidades pone en riesgo el sostenimiento de los sistemas públicos de salud.

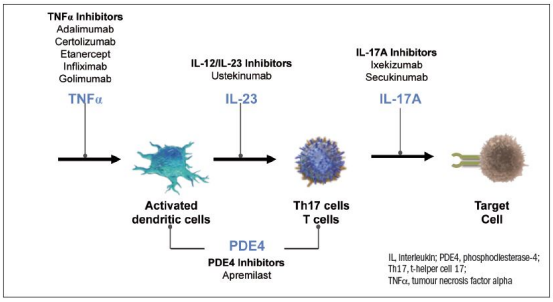


Figure 1. Pathogenesis-based approach to treatment of psoriasis and psoriatic arthritis. Adapted from Lynde et al. 2014[5]

Fig. 2 Dianas celulares y moleculares de los principales biológicos para el tratamiento de la psoriasis y la artritis psoriásica.

En este contexto, las decisiones terapéuticas deben estar basadas en el mejor conocimiento científico, respetando los valores y preferencias de los pacientes y empleando de forma eficiente los recursos. De este modo, podremos proporcionar una atención sanitaria eficaz al mayor número de personas[6]. Las **revisiones sistemáticas** constituyen el estándar científico que permite síntetizar los resultados obtenidos en los estudios primarios. Las revisiones sistemáticas de alta calidad son también base para el desarrollo de guías de práctica clínica sobre diagnóstico, pronóstico o intervenciones terapéuticas [7].

Las revisiones sistemáticas sintetizan las pruebas existentes en la literatura acerca de un problema específico clínico o de investigación empleando unos criterios metodológicos establecidos previamente. El hecho de que empleen una metodología estandarizada y unos protocolos de trabajo diseñados a priori constituye la clave para reducir la incertidumbre acerca de las conclusiones que derivan de ellos.

Los **meta-análisis** resultan de la aplicación de métodos estadísticos para resumir los resultados de estudios independientes obtenidos tras realizar una revisión sistemática. Permiten obtener estimaciones más precisas de la magnitud y dirección de los efectos potenciales de cada una de las alternativas evaluadas[8]. Los **meta-análisis en red** (NMA) realizan comparaciones directas e indirectas de más de dos alternativas entre sí y con el comparador común a todas ellas. Así pues, una revisión sistemática puede acompañarse de un meta-análisis, lo que dependerá, como veremos más adelante, de la naturaleza de las variables objetivo y de las condiciones del conjunto de estudios primarios sobre los que se trabaja.

Realizar una revisión sistemática es un proceso laborioso que requiere del conocimiento metodológico profundo y de una gran experiencia en el uso de los procedimientos y herramientas a emplear en cada una de las etapas de su desarrollo. La validez que lleguen a tener debería depender exclusivamente de la calidad metodológica de las mismas. la elaboración de guías para la conducción de revisiones sistemáticas [8–10] y de protocolos

para comunicarlas [64, 12, 13] está en constante desarrollo. Sin embargo, a pesar del cuidado con el que son conducidas, se han encontrado revisiones sistemáticas que responden de forma distinta a una misma pregunta de investigación. Debido a la repercusión de sus conclusiones resulta fundamental conocer la calidad metodológica y el riesgo de sesgo de dichos documentos para evaluar su aplicabilidad en la toma de decisiones. Aunque estos documentos de síntesis deben promover la mejora de los resultados en salud, en la actualidad se observa un auge en la producción de revisiones sistemáticas y meta-análisis de baja calidad científica[14].

La Cochrane define **calidad metodológica** como el rigor o la adherencia a los estándares más altos de investigación en la conducción de las revisiones sistemáticas. Y el riesgo de error sistemático es definido como **riesgo de sesgo**. La diferencia entre estos dos conceptos, que se solapan parcialmente, estriba en que el mero hecho de alcanzar los estándares mas altos de calidad (en especial algunos aspectos de los mismos como la aprobación de los estudios por comités de ética, el cálculo del tamaño muestral o la comunicación correcta de los resultados) no elimina completamente el riesgo de sesgo. Distinguir ambos conceptos ayuda a diferenciar entre la calidad de la comunicación de los resultados respecto a la calidad real de la investigación realizada [8]. En este sentido, en los últimos años se han desarrollado diferentes instrumentos para la evaluación tanto de la calidad metodológica como del riesgo de sesgo de las revisiones sistemáticas.

Actualmente existen diferentes grupos que trabajan en el desarrollo de revisiones sistemáticas o de guías de práctica clínica en el campo de la dermatología. El *Cochrane Skin Group* (CSG)[1] es un grupo de trabajo de Cochrane, organización internacional establecida en 1997 con el objetivo de producir documentos de síntesis basados en la evidencia para prevenir, diagnosticar y tratar enfermedades dermatológicas[15]. El alcance de su trabajo incluye desde revisiones sistemáticas relacionadas con melanoma hasta temas considerados de naturaleza cosmética.

Al identificar y priorizar los temas de las revisiones sistemáticas, el Cochrane Skin Group tiene en cuenta los siguientes criterios:

- El impacto de la condición en la vida de las personas, tanto psicológica como física.
- La carencia de conocimiento en el tratamiento de la afección motivo de análisis.
- La existencia de otras revisiones sistemáticas sobre el mismo tema.
- Lo que una revisión Cochrane podría aportar a lo que ya se conoce sobre un tema concreto.

[1] http://skin.cochrane.org/

- Realizar una evaluación de la calidad de las pruebas existentes sobre el tema propuesto.
- Analizar la importancia del tema a tratar para la salud pública.

Hasta la fecha, se han publicado más de 100 revisiones Cochrane que cubren diversos temas, desde la eficacia de las intervenciones hasta la seguridad de los fármacos. Nueve de estas revisiones tratan sobre distintos aspectos del tratamiento de la psoriasis. En las dos últimas décadas, el grupo CSG ha pasado a la vanguardia de la dermatología y la dermato-epidemiología basadas en la evidencia a escala internacional[15]. De hecho, las revisiones sistemáticas producidas por este grupo son de mayor calidad que las demás revisiones sistemáticas no-Cochrane publicadas sobre dermatología [16]. De lo anterior se deriva que la mayoría de las revisiones sistemáticas producidas por CSG ofrecen información realmente útil para mejorar la salud de los pacientes[17] y sirven de base para el desarrollo de nuevas guías de práctica clínica[18].

Incluso cuando las revisiones sistemáticas de CSG concluyen que son necesarios más estudios primarios, favorecen el desarrollo de nuevos ensayos clínicos a través de la *UK Dermatology Clinical Trials Network* (UK DCTN)[2][19]. CSG y UK DCTN se conectan con la *NLH Skin Disorders Specialist Library*[3] para difundir y poner en práctica dicho conocimiento.

Otras organizaciones que realizan revisiones sistemáticas son el *Joanna Brigss Institute*[4](JBI) o la Agencia de Evaluación de Tecnologías Sanitarias (AETSA) [5] del Sistema Andaluz de Salud. Además, existen diferentes asociaciones que transforman el conocimiento generado por las revisiones sistemáticas en recomendaciones clínicas en dermatología. En este sentido, el *European Dermatology Forum*, fundado en 1997 por un grupo de profesores europeos de dermatología, ha publicado más de 30 guías de práctica clínica[20]. Del mismo modo, la *British Academy of Dermatology* (BAD) desarrolla guías de práctica clínica que están acreditadas por el NHS Evidence[21]. Finalmente, la *American Academy of Dermatology* (AAD) ha publicado nueve guías de práctica clínica, seis de ellas relacionadas con la psoriasis.

[2]http://www.ukdctn.org/
[3]https://goo.gl/w1kXDP
[4]http://joannabriggs.org/
[5]http://www.aetsa.org/

II. Documentos de síntesis de la evidencia

II.a. Del arte de curar a la medicina basada en la evidencia

El arte de cuidar a los enfermos es tan antiguo como la humanidad, sin embargo la base científica de la medicina es reciente. El arte de curar se refiere a la actividad de la defensa de los pacientes por medio de la facultad humana y debe estar regido por tres principios, pues el médico es responsable del paciente en particular pero también de la sociedad en la que vive. Estos principios son:

- La primacía del bienestar del paciente sobre el conjunto de diagnósticos y opciones terapéuticas;
- La autonomía del paciente;
- La justicia social.

Se ha observado que en medicina, aunque basada en los fundamentos sólidos de las ciencias puras, los resultados obtenidos con el uso de procedimientos, instrumentos o medicamentos más o menos complejos, pueden variar entre los pacientes. Esto se ha demostrado así sobre todo cuando se toman decisiones basadas en la intuición, la experiencia no sistemática y el razonamiento fisiopatológico.

Por evidencia se entiende cualquier observación empírica sobre la aparente relación entre los eventos [22]. La idea de usar la evidencia en la práctica de la medicina estaba ya presente en la Grecia Antigua. Sin embargo el principio de la misma fue propuesto por Archibald L. (Archie) Cochrane en su libro *Effectiveness and Efficiency. Random Reflections on Health Services* donde se afirma que los recursos de atención de la salud son siempre limitados y deben ser [...]*eficientemente utilizados sobre la base de pruebas para proporcionar atención sanitaria eficaz al mayor número posible de personas.*[6]. En este sentido, desde 1992 se desarrolla un paradigma cuyo objetivo es consolidar la base científica de la medicina y reducir las incertidumbres existentes en la toma de decisiones[23]. Este paradigma ha dado lugar a los conceptos **medicina basada en la evidencia** y **salud pública basada en la evidencia** .

La definición de medicina basada en la evidencia ha evolucionado desde *el uso concienzudo y juicioso de la mejor evidencia actual de la investigación clínica en el manejo de pacientes individuales*[23] hasta *la integración de las mejores pruebas de investigación con la experiencia clínica y los valores de los pacientes*[24]. Estos valores desempeñan un papel importante, ya que representan los procesos que determinan lo que los pacientes y la sociedad ganan o pierden cuando se toma una decisión. La enumeración explícita y el equilibrio de los

beneficios y riesgos trae consigo juicios de valor subyacentes. La comprensión por los pacientes de este balance riesgo-beneficio ayuda a que sus preferencias y valores se muestren en sus decisiones. Así pues, de la medicina basada en la evidencia se ha pasado a la salud pública basada en la evidencia, que se define como la integración de las intervenciones basadas en la ciencia con las preferencias de la comunidad para mejorar la salud de la población [25].

No obstante, ambas concepciones se basan en el estudio de la evidencia para la redacción de guías en las que las opciones de actuación en salud se ordenan por su coste-efectividad. Esto permite desechar alternativas no eficaces, mejorar a largo plazo el estado de los pacientes, evitando complicaciones y tratamientos adicionales, e identificar medidas o promover leyes para la prevención de enfermedades. La creación de modelos estandarizados aumenta la homogeneidad en la toma de decisiones. Todo lo anterior tiene la intención de mejorar los resultados de cuidado, salud, costes y equidad.

Dada la diversidad de fuentes de observaciones empíricas que relacionan eventos es necesario establecer una jerarquía que ordene la validez de las mismas. Las jerarquías clasifican los estudios según el rigor de su metodología. Se representan en pirámides y si bien ninguna se adapta a todas las situaciones, son útiles para reflejar los conceptos de la medicina basada en la evidencia y para guiar respuestas a preguntas clínicas. .En la mayoría de las jerarquías las revisiones sistemáticas y los meta-análisis están en la parte superior de la pirámide y la opinión de expertos y la experiencia anecdótica en la parte inferior [26]. La mejora de las técnicas de evaluación de los estudios clínicos y la aparición de recursos prácticos que facilitan el acceso a la investigación de alta calidad ha dado lugar a un modelo de 6 niveles de evidencia. En la base se sitúan los estudios primarios y el punto más alto corresponde a los sistemas computerizados de toma de decisiones que generan evaluaciones o recomendaciones específicas del paciente. Las revisiones sistemáticas ocupan el tercer nivel en esta clasificación y consituyen la base de los niveles superiores [27]. Sin embargo el concepto inherente a todas estas pirámides, que la evidencia menos válida ,con mayor riesgo de sesgo, está en la base de las mismas no siempre es cierto. Por lo que deben ser guías pero no la regla en la toma de decisiones. La mayor limitación es que los niveles superiores dependen de los inferiores y si sólo se dispone de estudios de menor calidad las sinopsis y revisiones resultantes serán también de baja calidad.

Las revisiones sistemáticas tienen como objetivo reunir toda la evidencia que se corresponda con unos criterios de elegibilidad establecidos previamente para orientar una pregunta de investigación. Aplican métodos sistemáticos y explícitos para disminuir sesgos y aportar resultados fiables de los que se puedan extraer conclusiones y tomar decisiones [28]. Muchas se acompañan del meta-análisis que consiste en la aplicación de métodos

estadísticos con el fin de obtener estimaciones más precisas e investigar la consistencia y las diferencias entre estudios. El fin es que tanto los médicos como las autoridades sanitarias y los investigadores incorporen estos conocimientos a la toma de decisiones.

II.b. Elaboración de revisiones sistemáticas.

La ciencia de la síntesis de la investigación se encuentra en constante evolución. Existe una diferencia entre el estudio que se realiza y el informe que se publica por lo que se distinguen dos metodologías : conducción y reporte o notificación. Con respecto a la primera, diferentes instituciones desarrollan guías cuya intención es *"ayudar a tomar decisiones apropiadas en los métodos que se emplean en la revisión"* y cuyo espíritu es *"apoyar a ser sistemáticos, estar informados y ser explícitos pero no mecanicistas"*, en la formulación de las preguntas y el desarrollo de la revisión [8]. Además de la conducción su valor depende de cómo son reportadas [29] .

Metodología de la conducción de una revisión sistemática.

Emplearemos los conceptos desarrollados en el Manual de la Cochrane [8] centrándonos en las directrices MECIR (*Methodological Expectations of Cochrane Intervention Reviews*) [30]. De acuerdo a estos documentos, las principales fases para la elaboración de una revisión sistemática son:

- **Elaboración de un protocolo *a priori*;**
- **Formulación de la pregunta de investigación y definición a priori de los objetivos y criterios de inclusión;**
- **Selección de los estudios y extracción de los datos;**
- **Evaluación del riesgo de sesgo de los estudios incluidos.**

Elaboración de un protocolo *a priori*. La publicación de un protocolo *a priori* reduce el impacto de los sesgos, estimula la transparencia de los métodos y los procesos planteados, disminuye la posibilidad de duplicación, y permite la revisión por pares de los métodos propuestos [31]. Existen diferentes repositorios donde pueden registrarse protocolos cuyo cometido es realizar un seguimiento prospectivo de su cumplimiento como, por ejemplo, *The International Prospective Register of Systematic Reviews* (PROSPERO); , el primer registro internacional de protocolos de revisiones sistemáticas abierto en 2011, o *The Joanna Briggs Institute.* Los protocolos también pueden publicarse en revistas como Systematic Reviews [6], la primera revista internacional nacida en 2012 y dedicada a publicar exclusivamente revisiones sistemáticas, incluídos sus protocolos. En cualquier

[5] https://www.crd.york.ac.uk/prospero/
[5] http://joannabriggs.org/research/registeredtitles.aspx
[6] https://systematicreviewsjournal.biomedcentral.com/

caso, siempre debería estar disponible por si es requerido por los revisores o editores de las revistas científicas.

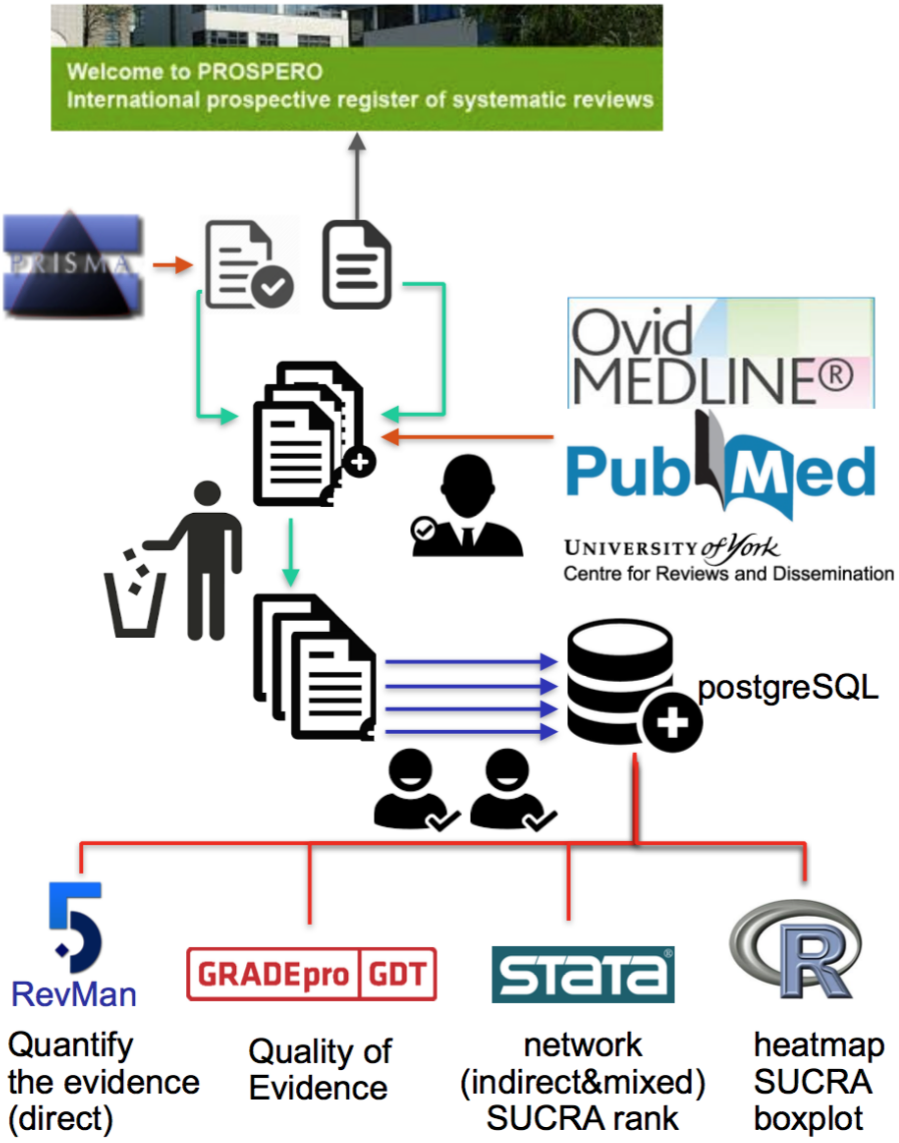


Fig. 3 Flujo de trabajo que representa las diferentes fuentes de datos, etapas y herramientas empleadas para el desarrollo de una revisión sistemática y meta-análisis en red.

Los autores deben hacer el esfuerzo de mantenerse fieles al protocolo publicado porque las decisiones a posteriori son muy susceptibles de sesgos. Sin embargo, en ocasiones es apropiado realizar cambios cuando los métodos de un asunto específico no se han incluído en el protocolo, no se pudieron emplear los descritos o se siguió una alternativa preferible. Los

cambios introducidos deben ser explicitados. En ningún caso las alteraciones del protocolo inicial deben basarse en los resultados obtenidos tras llevar a cabo la investigación.

UNIVERSITY of York
Centre for Reviews and Dissemination

NHS
National Institute for Health Research

PROSPERO International prospective register of systematic reviews

Quality assessment of pharmacological and non-pharmacological interventions for the management of psoriasis: a systematic review of reviews and network meta-analysis

Juan Ruano, Francisco Gomez-Garcia, Beatriz Maestre-López, Beatriz Isla-Tejera, Marcelino González-Padilla, Jesus Gay-Mimbrera, Macarena Aguilar-Luque, Pedro Carmona, Antonio Vélez García-Nieto, Juan Luis Sanz-Cabanillas

Citation
Juan Ruano, Francisco Gomez-Garcia, Beatriz Maestre-López, Beatriz Isla-Tejera, Marcelino González-Padilla, Jesus Gay-Mimbrera, Macarena Aguilar-Luque, Pedro Carmona, Antonio Vélez García-Nieto, Juan Luis Sanz-Cabanillas. Quality assessment of pharmacological and non-pharmacological interventions for the management of psoriasis: a systematic review of reviews and network meta-analysis. PROSPERO 2016:CRD42016041611 Available from http://www.crd.york.ac.uk/PROSPERO_REBRANDING/display_record.asp?ID=CRD42016041611

Review question(s)
To investigate the methodological quality of systematic reviews addressing the efficacy and/or safety of pharmacological interventions (biologics, systemics, topical) for chronic plaque psoriasis in adult patients.

To investigate the methodological quality of systematic reviews addressing the efficacy and/or safety of non-pharmacological interventions (phototherapy, skin moistures, diet) for chronic plaque psoriasis in adult patients.

To investigate the quality of reporting of systematic reviews addressing the efficacy and/or safety of pharmacological interventions (biologics, systemics, topical) for chronic plaque psoriasis in pediatric patients.

To investigate the quality of reporting of systematic reviews addressing the efficacy and/or safety of non-pharmacological interventions (phototherapy, skin moistures, diet) for chronic plaque psoriasis in pediatric patients.

To compare the quality of reporting of systematic reviews addressing the efficacy and/or safety of interventional (pharmacological and non-pharmacological) vs observational studies (pharmacogenetics, co-morbidities, etc) for chronic plaque psoriasis in adult or pediatric patients.

To explore if bibliometric factors (journal and researchers) can explain differences in the quality of reporting of systematic reviews in chronic plaque psoriasis.

Fig. 4 Ejemplo de protocolo publicado *a priori* en PROSPERO.

Formulación de la pregunta de investigación y desarrollo de los criterios de inclusión de los estudios. La correcta formulación de la pregunta de investigación determina la validez externa de la revisiones sistemáticas y guía las siguientes fases del proceso de revisión. Se han desarrollado diferentes estrategias en función de la disciplina y del tipo de investigación [32]. PICO (Participantes; Intervención;Comparador y Objetivos/resultados) es la más empleada en la formulación de preguntas clínicas [33].

Participantes (P): Los criterios deben ser lo suficientemente amplios para abarcar la heterogeneidad de estudios y permitir obtener una respuesta significativa cuando sean agrupados. Cualquier restricción debe estar basada en la evidencia preexistente y documentada para evitar sesgos por arbitrariedad de los criterios. Se aconsejan dos pasos en su determinación:

1. *Definir explícitamente la condición o enfermedad, evitando la exclusión de algunos estudios por cambios recientes en los criterios o métodos diagnósticos o terapéuticos de una enfermedad;*

2. *Identificar correctamente la población y el ámbito de interés: hay que definir bien edad, género, raza, nivel educativo , condición especial y ámbito: población general, comunidad cerrada, hospitalaria, etc.*

Intervención/Comparador (I/C): Puede tratarse de fármacos o de intervenciones más complejas como las educacionales o comportamentales. Cuando se trata de fármacos deben considerarse desde la preparación o vía de administración hasta la frecuencia, dosis o duración. En las segundas deben definirse todos los aspectos centrales de las mismas.

Objetivos (O): Deberían incluirse todos los objetivos relevantes para la toma de decisiones. Hay que tener en cuenta la literatura publicada al respecto, la experiencia de autores y asesores, así como las preferencias de los pacientes. No tienen necesariamente que haber sido recogidos en los estudios primarios. Deben incorporarse objetivos relacionados con aspectos favorables, desfavorables y económicos.

Los objetivos relevantes pueden diferenciarse en:

- **Principales**: deben guiar la búsqueda de los estudios primarios. Se aconseja no definir más de tres.

- **Secundarios**: están formados por el resto de objetivos relevantes mas aquellos que ayudan a explicar el efecto del fármaco.

Para cada objetivo, debe especificarse:

1. El **tipo de escala**, indicando si ha sido publicada y validada;

2. El **momento de la medida**. Una estrategia es agrupar el tiempo en intervalos pre-especificados (p.e.: “corto”, “medio” y “largo“ plazo);

3. La **medida del efecto resumen**, que puede diferir entre los estudios incluidos:

    - Para los **resultados binarios**, las más comunes son: la razón de riesgo (odds ratio, OR) y la diferencia de riesgo [34]. Aunque los efectos relativos son más consistentes entre los estudios que los efectos absolutos [35], hay que tener en cuenta que las diferencias absolutas son importantes para interpretar desde un punto de vista clínico los hallazgos.

- Para los **resultados continuos** suele emplearse la diferencia de las medias, cuando los estudios emplean la misma escala, o la diferencia estandarizada de las medias en caso contrario.

A lo anterior puede añadirse el *tipo de estudio* elegible *(S)*. *Tipo de estudio*: La aleatorización es la única manera de prevenir diferencias sistemáticas entre las características basales de los participantes. La evidencia sugiere que los ensayos clínicos no aleatorizados producen estimaciones de efecto con beneficios más extremos que los ensayos clínicos aleatorizados.

El **alcance de la pregunta de investigación** puede ser amplio o restringido. El primero permite un resumen más amplio de la evidencia pero tiene mayor riesgo de heterogeneidad. El enfoque restringido es de más fácil manejo pero conlleva una menor capacidad de extrapolación de la evidencia a situaciones ajenas a las condiciones de búsqueda establecidas. Es posible realizar **cambios en la formulación de la pregunta de investigación**, siempre que ayuden a explorar temas inesperados que surjan durante el proceso de búsqueda y extracción de los datos [36]. Pero no en fases más avanzadas del proceso y menos si están dirigidos por los resultados finales. Finalmente, cualquier modificación debe quedar especificada en el protocolo y en la revisión sistemática en una sección denominada "Diferencias entre el protocolo y la revisión sistemática".

Selección de los estudios y extracción de los datos. El objetivo final de la revisión sistemática es incorporar todos los estudios relevantes, por lo tanto, la búsqueda debe ser amplia y objetiva. Todo el proceso debe ser descrito en el protocolo y en la sección Métodos de la revisión sistemática. Los resultados obtenidos se reflejarán en el texto y serán presentados mediante un diagrama de flujo en una figura con el objetivo de favorecer la transparencia y reproducibilidad. Es aconsejable que un experto en biblioteconomía participe en el diseño de la estrategia de búsqueda.

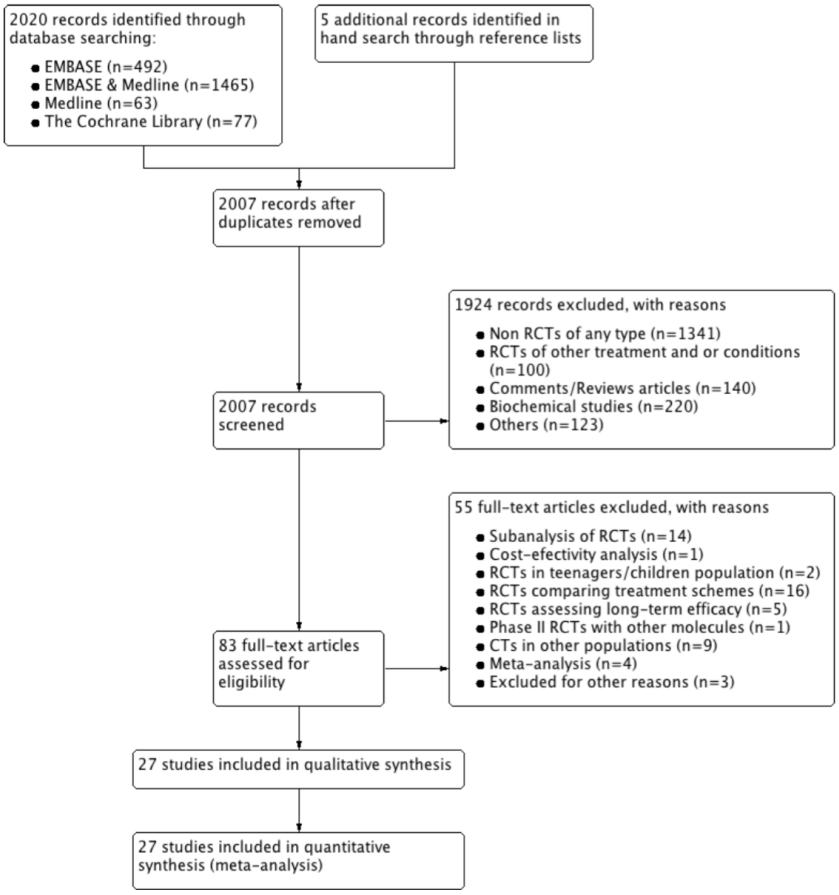


Fig. 5 Ejemplo de diagrama de flujo para la representación del proceso de evaluación y selección de los estudios de una revisión sistemática.

Estrategia de búsqueda. Debe guiarse por los criterios de elegibilidad y emplear los términos de búsqueda derivados de los elementos PICO. Es necesario mantener un equilibrio entre la sensibilidad y la precisión, teniendo en cuenta que el tiempo y el presupuesto limitan esta etapa de la revisión sistemática.

A continuación se exponen algunos principios para la estrategia de búsqueda [37]:

- **No confiar sólo en los filtros de búsqueda.**. La mayoría de estos filtros están disponibles para MEDLINE [38] y EMBASE[39] y funcionan mejor cuando se evalúan que cuando se usan en el contexto real de una revisión sistemática.
- **No confiar exclusivamente en el vocabulario controlado.** La publicación y la indexación de los estudios es muy variable por lo que no suelen emplearse todos los términos del vocabulario. El uso de palabras de texto para pruebas médicas particulares ayuda a identificar los artículos que aún no han sido indexados o los

que lo han sido de forma incorrecta.[38] En cualquier caso, las restricciones de lenguaje, fecha y formato del documento deben estar bien justificadas y notificadas.

- **Llevar a cabo la búsqueda en múltiples fuentes.** La combinación de búsquedas altamente sensibles con fuentes adicionales de información como el seguimiento de citas, la lectura de referencias y la identificación de artículos que citan estudios clave son las mejores formas de encontrar citas adicionales [40].

Fuentes de información. Las bases de datos bibliográficas pueden ser clasificadas según la tabla 1. Debe buscarse en al menos dos bases de datos de este tipo. Las tres más usadas para encontrar ensayos clínicos aleatorizados son MEDLINE, EMBASE y el Registro Central Cochrane de Ensayos Controlados (CENTRAL). Existe cierta superposición de documentos entre EMBASE y MEDLINE, por lo que pueden consultarse ambas bases de datos de forma simultánea desde la URL de EMBASE.com que incluye registros de MEDLINE desde 1966. Puede complementarse esta búsqueda con las bases de datos nacionales, regionales y de temas específicos. el Índice de Citas Científicas/Índice de Citas Científicas Ampliado(SciSearch/Red de la Ciencia) enumera los artículos y los enlaza con los que los han citado. Además, Scopus de Elsevier incluye revistas y actas de conferencias. Acerca de las bases de datos de literatura gris, existe evidencia de que la significación de los resultados y la magnitud del efecto es menor en los documentos obtenidos de estas fuentes que en los estudios publicados en revistas científicas.

Table 1 Tipos de bases de datos

Generales	MEDLINE, EMBASE y Registro Central Cochrane de Ensayos Controlados (CENTRAL).
Nacionales y regionales	Índice Médico para la Región Oriental del Mediterráneo, PASCAL (Europa), IndMED (India), Korea med (Korea), LILACS (Latinoamérica y El Caribe), Panteleimon (Ucrania y Federación Rusa), etc.
De temas específicos	De promoción de la salud (BiblioMap-EPPI Base); de salud internacional (POPLINE); de Enfermería y Ciencias de la Salud (EMCare); de Psiquiatría (PsycINFO), etc.
Índice de citas	Para revistas científicas, técnicas y médicas: El Indice de Citas Científicas/Índice de Citas Científicas Ampliado (SciSearch/Red de la Ciencia).

Bases de datos de tesinas y tesis	ProQuest ,Índice de Tesis de Gran Bretaña e Irlanda, DissOnline (Alemania).
Bases de datos de literatura gris	SIGLE (sistema de información de la literatura gris) en Europa-Francia, HMIC (base de datos del Consorcio de Manejo de Información Sanitaria) en Inglaterra, NTIS (Servicio Nacional de Información Técnica) en Estados Unidos, etc.

Revistas y otras fuentes que no son bases de datos bibliográficas. Existen diversas estrategias de encontrar información adicional a la obtenida mediante las búsquedas en bases de datos previamente mencionadas.

- **Búsqueda Manual.** La evidencia ha demostrado que es necesario emplear la búsqueda manual porque no todos los informes de los ensayos se incluyen en las bases de datos bibliográficas ni contienen los términos de búsqueda específicos.

- **Revistas con texto completo disponibles electrónicamente.** Permiten el acceso a partes de la revista no disponibles en la edición física. Es importante especificar que el texto completo consultado es electrónico porque las condiciones pueden cambiar (pérdida del acceso, disponibilidad). Se aconseja hacer copia.Vg, Central BioMed, Biblioteca Pública de la Ciencia (PLOS), Central Pubmed (PMC), etc...

- **Resúmenes y Actas de Conferencias.** Cerca de la mitad de los ensayos clínicos que se informan de este modo nunca llegan a publicarse completos y los que lo hacen son sistemáticamente diferentes.Pueden estar disponibles en material impreso, en CD-ROM en internet o en suplementos de revistas como BIOSIS.[7]

- **Otras revisiones sistemáticas , guías y listas de referencia como fuentes de estudios.** Pueden encontrarse en la Base de datos de la Cochrane para Revisiones Sistemáticas, La Base de Datos de Resúmenes de Revisiones de Efectos (DARE), o la Base de Datos de Evaluación de Tecnología en Salud (HTA). Ejemplos de guías basadas en la evidencia son: NICE (Reino Unido), Grupo de Guías de Nueva Zelanda o Guía Nacional para el intercambio de información (EEUU).

- **Búsqueda en la web.** Existe poca evidencia sobre la rentabilidad de esta estrategia. Pueden ser de interés: sitios web de fondos de investigación , fabricantes o industria farmacéutica. Se recomienda guardar copia de la información y registro de la fecha de consulta.

[7] http://www.biosis.org/

- **Estudios no publicados y en proceso.** Estos registros tienen un valor especial porque algunos resultados de ensayos o no se publican o lo hacen de forma parcial. Existen iniciativas que ayudan en este sentido como: 1) el Número de Registro Estándar Internacional del Ensayo ClÍnico Controlado Aleatorizado que asigna números únicos de ensayos clínicos aleatorizados a nivel internacional [8]; 2) el apoyo al registro de los ensayos por parte de los editores de las revistas o la Política de Acceso Público de los Institutos Nacionales de Salud [9]; 3) La Plataforma Internacional de Registro de Ensayos clínicos creada por la Organización Mundial de la Salud. Otra forma de acceso a los datos son las cartas formales de solicitud de información.

Selección de los estudios y extracción de los datos.

Selección de los estudios. Todo el proceso debe ser informado en la parte de Métodos.Se recomienda utilizar formularios pilotados y asegurar la consistencia en la aplicación de los criterios de elegibilidad. La participación de dos investigadores puede reducir la probabilidad de error de rechazar informes pertinentes [41] . Los expertos sobre el tema pueden tener opiniones preestablecidas por lo que es de valor que un revisor no lo sea. El beneficio puede ser mayor para los temas en los que la selección requiere juicios difíciles [42]. En estos casos los autores deberían informar sobre el nivel de acuerdo entre los evaluadores (en general no se aconseja), la frecuencia de arbitraje sobre la selección y los esfuerzos realizados para resolver los desacuerdos (discusión entre revisores, participación de una tercera persona). Deben citarse los informes excluidos con el motivo de la exclusión. Finalmente, un estudio se puede haber publicado en más de un informe lo que puede dar lugar a error si los resultados de los estudios se incluyen más de una vez en el meta-análisis.

Extracción de los datos. Se entiende por dato cualquier información contenida en un estudio. Es necesario planificar por adelantado **qué datos se buscarán** , desarrollar una estrategia para recuperarlos y realizar una prueba piloto con el formulario de extracción de datos empleando una muestra representativa de los estudios. Existe evidencia indirecta de que la extracción independiente por **al menos dos revisores** genera menos errores. El desacuerdo entre investigadores puede ser resuelto mediante discusión, intervención de una tercera persona o contactando con los autores del estudio. Estos desacuerdos deben quedar anotados y el grado de discrepancia entre evaluadores para cada estudio primario y/o item puede ser calculado mediante la obtención del estadístico kappa. Se aconseja que un revisor sea metodólogo y otro un experto sobre el tema.

Con respecto al **formulario estandarizado de obtención de datos** no existe evidencia empírica de que un formato electrónico sea superior a uno analógico. Sus funciones son:

[8] https://clinicaltrials.gov/
[9] https://publicaccess.nih.gov/

- Es la fuente de datos para la inclusión en el análisis;
- Proporciona un resumen de la pregunta investigación y los criterios de elegibilidad;
- Es el registro histórico de las decisiones durante el proceso de revisión.

Evaluación del riesgo de sesgo de los estudios incluidos.

Definición de sesgo. Se define sesgo como el error sistemático en los resultados o inferencias. Los sesgos pueden producirse en cualquier dirección y son variables en su magnitud. Generalmente no es posible conocer hasta qué punto los sesgos afectan los resultados de un estudio. Por este motivo es más apropiado hablar de riesgo de sesgo[8]. Las diferencias en los riesgos de sesgo ayudan a explicar la heterogeneidad de resultados entre los estudios siendo más probable que los más rigurosos estén más cerca de la verdad.La validez interna viene dada por la ausencia de errores sistemáticos y representa el riesgo de que un estudio sobreestime o subestime el verdadero efecto de la intervención o método diagnóstico o preventivo. Se mide mediante la evaluación del riesgo de sesgo de cada uno de los estudios de la revisión sistemática.

El **riesgo de sesgo** debe diferenciarse de la **calidad metodológica** que se corresponde con el grado de cumplimiento con los estándares más altos en la conducción del estudio. Es posible realizar un estudio con alta calidad metodológica y que presente un alto riesgo de sesgo (Vg. si no es posible el cegamiento). Además, algunos estándares de alta calidad como el cálculo del tamaño muestral o la aprobación del estudio en comité de ética es poco probable que afecten al riesgo de sesgo.En la tabla 1. se ofrecen los criterios generales de evaluación del riesgo de sesgo.

Dado que la mayoría de revisiones sistemáticas realizadas hasta el momento se han centrado en los ensayos clínicos aleatorizados, repasaremos las fuentes de sesgo de este tipo de estudios así como las herramientas empleadas para su evaluación.

Fuentes de sesgo de los ensayos clinícos aleatorizados. Para todas las fuentes de sesgo es importante considerar la magnitud y dirección probable del sesgo. Atendiendo a la fuente del sesgo, se consideran los siguientes tipos:

- **Sesgo de selección.** Hace referencia a las diferencias sistemáticas entre las características iniciales de los grupos que se comparan. La evidencia ha demostrado que es necesario interrelacionar dos procesos para controlarlo:
  - **Generar una secuencia de asignación aleatorizada** para evitar efectos de intervención sesgados con estimaciones exageradas.

  - **Ocultar la secuencia de asignación**, ya que el conocimiento de la siguiente asignación conlleva el riesgo de reclutamiento selectivo de los participantes dando lugar a una posible sobreestimación del efecto.

- **Sesgo de realización.** Se refiere a las diferencias sistemáticas en la asistencia o en la exposición a otros factores. El control se realiza mediante el cegamiento de los participantes y el personal. En aquellos estudios diseñados sin ciego simple o doble, se aprecia una sobreestimación del efecto de la intervención, que es mayor cuando las variables evaluadas son de naturaleza subjetiva.

- **Sesgo de detección.** Hace referencia a la diferencias sistemáticas en la forma de obtener los resultados. Es necesario tener en cuenta: quién evalúa el resultado y la naturaleza del resultado medido (objetivo o subjetivo). El cegamiento de los evaluadores es imposible en algunos estudios, como aquellos que evalúan procedimientos quirúrgicos, por lo que se deben emplear protocolos estrictos para evitar diferencias sistemáticas en la forma de tratar a los pacientes. También puede ser difícil evitar este tipo de sesgo en estudios llevados a cabo con fármacos de gran eficacia o con efectos adversos selectivos, respecto al comparador.

- **Sesgo de desgaste.** Se refiere a las diferencias sistemáticas entre los grupos a comparar en relación a la tasa de abandono del estudio. Existen distintas formas de analizar los resultados teniendo en cuenta este tipo de sesgo. De hecho, existe evidencia de una sobreestimación a favor de la intervención cuando se excluye participantes o cuando se realiza un análisis por protocolo.

  - **Análisis por intención de tratar (ITT).** Es la forma de tratar los datos que conlleva menor riesgo de sesgo. Se mantienen a los pacientes en los grupos asignados independientemente de la intervención recibida , se miden todos los datos y se incluye a todos los pacientes en el análisis. Es frecuente encontrar análisis por ITT modificada dada la alta frecuencia de datos faltantes.
  - **Análisis por protocolo (APP).** Deben tratarse como de alto riesgo de sesgo dado que los cambios de grupo de tratamiento puede estar relacionados con el pronóstico.
  - **Imputaciones de datos perdidos.** Se utilizan para paliar los datos faltantes y tratarlos como si fueran reales. Puede originar sesgos graves e intervalos de confianza estrechos.Puede emplearse el desenlace medio o la última observación realizada.

- **Sesgo de notificación.** Se refiere a las diferencias sistemáticas entre los resultados presentados y los no presentados. Es más probable que se notifiquen los análisis con diferencias estadísticamente significativas. Las causas más frecuentes de no publicar un desenlace son: "falta de importancia clínica" o "falta de significación estadística" Justificación: Existe evidencia de que los desenlaces estadísticamente significativos tienen más probabilidades de estar descritos para los datos de eficacia y esto actúa a favor de la intervención.
- **Otros sesgos.** Pueden ser relevantes sólo en determinadas circunstancias dependiendo del tipo de ensayo (arrastre, cruzados) o del ámbito clínico determinado.

Herramientas de evaluación del riesgo de sesgo.

Existen diferentes herramientas desarrolladas para la medición del riesgo de sesgo. Dado que es imposible conocer el riesgo de sesgo real de un estudio, la posibilidad de validación de las herramientas es muy limitada. La mayoría de los instrumentos desarrollados son escalas o listas de verificación que ofrecen sistemas de puntuación parciales o totales. Sin embargo, existen pruebas empíricas que desaconsejan su uso ya que son poco fiables y transparentes y resulta difícil justificar la ponderación que hacen de los diferentes ítems. Un ejemplo es la herramienta desarrollada por Jadad que se desaconseja en la actualidad [43].

La tendencia en este momento es emplear instrumentos estructurados en dominios en los que se evalúan componentes específicos de las revisiones sistemáticas. Los distintos niveles de formación metodológica pueden dan lugar a evaluaciones diferentes. Es importante que sea llevada a cabo tanto por metodólogos como por expertos en el tema revisado.

Una dificultad asociada a este proceso es la notificación incompleta de los métodos y resultados obtenidos, lo que puede hacer necesario que los evaluadores traten de contactar con los autores de los estudios para solicitarles dichos datos o alguna aclaración sobre los mismos.

Se aconseja emplear la **Herramienta de la Cochrane de evaluación del riesgo de sesgo** [44] cuya evidencia ha sido demostrada y que consta de 7 dominios:

- Aleatorización.
- Ocultación de la aleatorización.
- Cegamiento de los participantes e investigadores.
- Cegamiento de los evaluadores.
- Sesgo de desgaste.

- Sesgo de reporte.
- Otros sesgos en función del ensayo clínico.

La **Herramienta de la Cochrane de evaluación del riesgo de sesgo** permite clasificar el riesgo de sesgo de cada estudio y entre los estudios en:

- Bajo riesgo de sesgo.
- Riesgo de sesgo incierto.
- Alto riesgo de sesgo.

Las evaluaciones resumidas del riesgo de sesgo para un resultado dentro de cada ensayo deben influir en los procedimientos siguientes de análisis . En función de dichas evaluaciones pueden definirse diferentes estrategias para el meta-análisis que se exponen más adelante.

La presentación del **resumen del riesgo de sesgo** debe incluir [8]:

- Dirección probable del sesgo.
- Magnitud probable del sesgo.

La base de la evidencia empírica no proporciona todavía información clara sobre las situaciones particulares en las cuales los sesgos pueden ser grandes o pequeños. Sin embargo, es posible considerar la magnitud probable del sesgo con relación a la estimación de la magnitud del efecto.

El resumen del riesgo de sesgo se puede presentar en cuatro niveles:

- A través de sus desenlaces.
- De un desenlace dentro de un estudio.
- De un desenlace a través de todos los estudios.
- Para una revisión general.

II.c. Síntesis de los resultados de revisiones sistemáticas.

El enfoque dado al proceso de síntesis de los resultados tras la búsqueda sistemática, depende de la naturaleza de la pregunta de revisión y de los estudios primarios que se están sintetizando.

La **síntesis narrativa** emplea métodos subjetivos. Se debe preespecificar, justificar y seguir sistemáticamente el método para no hacer énfasis en los resultados de un estudio sobre los de otros.

Mediante el **meta-análisis** se analizan los datos cuantitativos. Su uso aporta valor a la revisión realizada porque permite aumentar la potencia estadística, la precisión, responder a preguntas no planteadas en estudios individuales, resolver controversias que surgen de estudios aparentemente contradictorios y generar nuevas hipótesis. Sin embargo, el mero hecho de emplearlo no garantiza que los resultados de la revisión sean más válidos que los de los estudios individuales dado que puede utilizarse de forma inapropiada[8].

En ambos casos, los aspectos más importantes a considerar en la **síntesis de los resultados** son:

- Si el enfoque analítico es apropiado para la pregunta de investigación planteada.
- Si se tiene en cuenta la variación entre estudios (heterogeneidad).
- Si se tienen en cuenta los sesgos en los estudios primarios.
- Si se ha completado la información de los estudios.
- Si los revisores han introducido sesgos en la forma en que informan sus hallazgos.

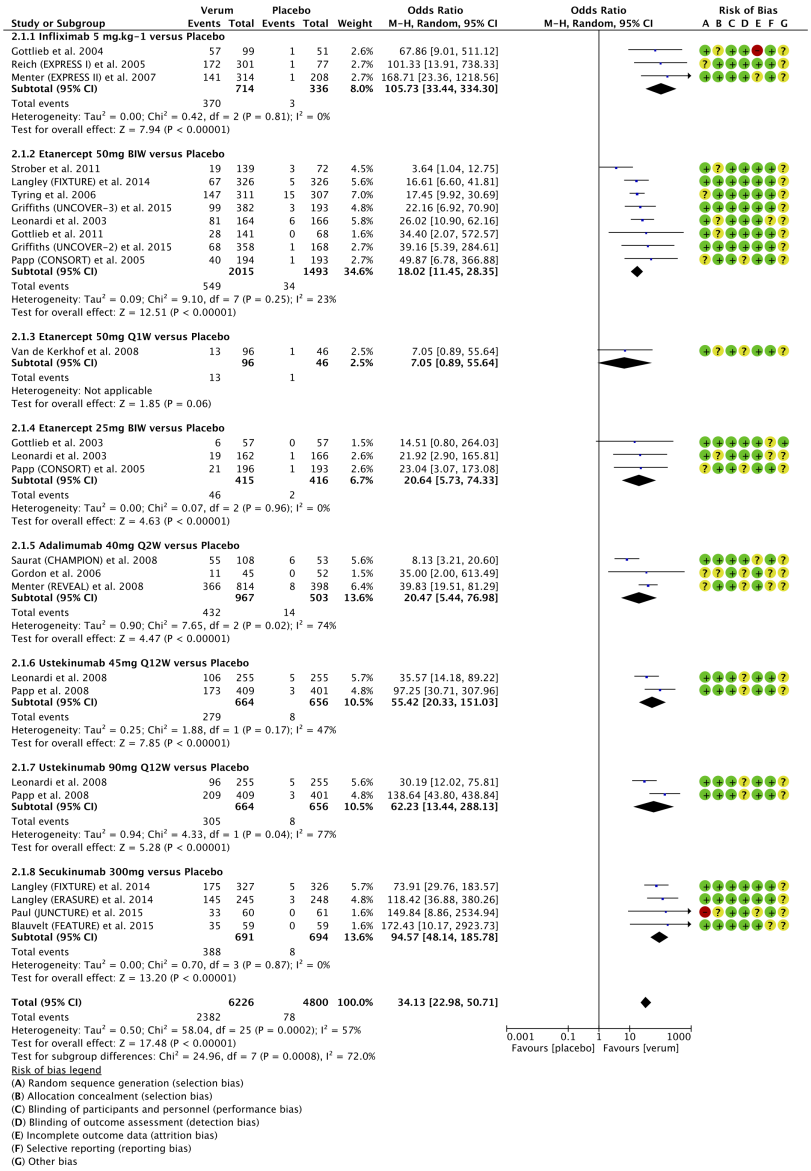


Fig. 6 Forest plots.

Síntesis cuantitativa

La forma de medir un efecto depende del tipo de dato disponible. Para los estudios de efectividad se consideran cinco tipos de datos:

- Dicotómicos. Cuya medidas de efecto son: razón de riesgos (RR o riesgo relativo), *odds ratio* (OR), diferencia de riesgos (RA o reducción del riesgo absoluto) y número necesario a tratar (NNT).

- Continuos. Cuyas medidas de efecto más frecuentemente empleadas son la diferencia de medias y la diferencia de medias estandarizada.
- Ordinales. Cuyas medidas de efecto más empleadas son las escalas de medición de las que es necesario tener en cuenta que hayan sido validadas o los OR proporcionales.
- Recuentos y tasas de recuentos. Emplean los datos de recuento que pueden dividirse en poco frecuentes o muy frecuentes. También es muy empleado el cociente de tasas.
- Tiempo hasta el suceso. Emplean los datos de supervivencia. La forma más apropiada en este caso es el cociente de riesgos instantáneos (HR o hazard ratio).

Para la síntesis de los datos se pueden emplear, de forma general, **modelos de efectos fijos**, si se supone que cada estudio estima el mismo efecto de intervención, o **modelos de efectos aleatorios**, en caso contrario. Sin embargo, hay muchas variantes y extensiones con las opciones de modelar los datos de resultados explícitamente:

- Con un enfoque de regresión logística para los datos binarios[45].
- Un meta-análisis bivariado o multivariante, cuando se evalúan dos o más resultados simultáneamente[46].
- Meta-análisis en red, cuando se comparan efectos múltiples, directos e indirectos, de diferentes opciones de tratamiento [47].
- Metaregresión, cuando se quiere modelar la variación en los efectos del tratamiento[48].

En otros tipos de preguntas de investigación como la precisión de prueba de diagnóstico, un enfoque bivariado se ha convertido en el método estándar; en éste, la sensibilidad y la especificidad son modelados simultáneamente para tener en cuenta su correlación[49].

Análisis de heterogeneidad

Se denomina heterogeneidad a cualquier tipo de variabilidad de los estudios que integran una revisión sistemática. La heterogeneidad puede ser:

- **Clínica**, por diferencias entre participantes, intervenciones o resultados.
- **Metodológica**, por diferencias en el diseño y riesgo de sesgo de los estudios incluidos.
- **Estadística**, consecuencia de las anteriores, se manifiesta en que los efectos de la intervención difieren entre sí más de lo que se esperaría si se debieran solo al error aleatorio.

El meta-análisis se debería realizar sólo cuando el grupo de estudios que han sido seleccionados para sintetizar los resultados de cada variable fuera lo suficientemente homogéneo como para proporcionar un resumen significativo. Para ello es importante evaluar la heterogeneidad antes de llevar a cabo la síntesis de los resultados.

Un signo indirecto de heterogenidad es que los intervalos de confianza de los resultados de los estudios individuales muestran escasa superposición. La prueba estadística de **Ji cuadrado (Chi2)** permite evalúar si las diferencias observadas en los resultados son compatibles con el azar. Esta prueba tiene una potencia estadística baja en aquellos meta-análisis con pocos estudios primarios o que fueron desarrollados con tamaños muestrales pequeños. Un resultado no significativo por lo tanto, no se debe tomar como prueba de falta de heterogeneidad. Además cuando hay muchos estudios en un meta-análisis, la prueba tiene una potencia estadística grande para detectar una pequeña cantidad de heterogeneidad que pudiera no ser clínicamente importante.

Como siempre existe diversidad clínica y metodológica , la heterogeneidad estadística es inevitable y existirá se tenga o no la capacidad de detectarla. Por ello se han desarrollado métodos para cuantificar la inconsistencia que evalúan el impacto de la hetereogenidad en el meta-análisis. Un estadístico útil para cuantificar la inconsistencia es I^2, que describe el porcentaje de la variabilidad de las estimaciones del efecto que es debido a la heterogeneidad en lugar del azar[50]. Se han establecido umbrales para la interpretación de los valores de I^2. Así, se considera que la heterogeneidad es elevada cuando el estadístico se encuentra entre el 75% y el 100%.

Si la heterogeneidad de un grupo de estudios es significativa, se puede optar por varias opciones:

- Verificar si los datos son correctos, ya que puede haber habido un error en la extracción de los datos de las fuentes primarias o en la incorporación de los mismos en el proceso de análisis cuantitativo.
- Citar un valor promedio del efecto de la intervención y no realizar el meta-análisis.
- Explorar la heterogeneidad mediante un análisis de subgrupos más homogéneos o realizar una metaregresión. En este caso debe haber sido preespecificado en el protocolo redactado *a priori*.
- Ignorar la hetereogeneidad, como en el meta-análisis de efectos fijos.
- Realizar un meta-análisis de efectos aleatorios.

- Cambiar la medida de efecto.
- Excluir los estudios que introducen la heterogeneidad aunque esto último no es recomendable.

Análisis de sensibilidad

Realizar un análisis de sensibilidad disminuye la incertidumbre acerca de los resultados obtenidos. Si se demuestran incertidumbres importantes pueden tratar de resolverse contactando con los autores de los estudios primarios u obteniendo datos de pacientes individuales. Si no es posible, los resultados deben interpretarse con precaución antes de llegar a conclusiones finales y teniendo en cuenta que, en cualquier caso, dichos hallazgos permiten generar nuevas hipótesis de investigación.

Análisis del sesgo de publicación

El sesgo de publicación surge cuando la diseminación de los hallazgos de la investigación está condicionada por la naturaleza y la dirección de los resultados obtenidos. En la tabla 2 se exponen diferentes tipos de sesgo de publicación.

Se han descrito dos causas que pueden explicar la existencia de este tipo de sesgo. En primer lugar, los estudios con resultados negativos permanecerían sin ser publicados porque los autores no redactarían el texto, los revisores por pares serían menos favorables a su evaluación positiva y los editores no aceptarían publicarlos en las revistas de investigación. Por otro lado, el tipo de financiación que da soporte al desarrollo de los estudios podría condicionar que los resultados fueran comunicados a la comunidad científica. De hecho, la financiación pública se asocia significativamente con mayor probabilidad de publicación [51] en comparación con la de la industria farmacéutica [52]. En este sentido, es más probable que los estudios patrocinados por la industria farmacéutica tengan resultados favorables en comparación con los financiados por otras entidades como instituciones académicas[53].

Existen dos formas de reducir el sesgo de publicación cuando realizamos una revisión sistemática:

- **Inclusión de estudios no publicados en las revisiones sistemáticas.** Se ha observado que los ensayos clínicos publicados tienen un mayor efecto de la intervención que los no publicados[54]. Sin embargo, es posible que los informes de estudios incluidos obtenidos a partir de la literatura gris no sean representativos de los estudios no publicados. Además, podrían tener una menor calidad metodológica y resultar más difícil el contacto con sus autores.

- **Emplear registros de ensayos clínicos.** Dichos registros deben posibilitar la búsqueda electrónica, ser de libre acceso para su consulta y registro, estando administrados

por una organización sin ánimo de lucro. Desde 2005 el *International Committee of Medical Journal Editors* (ICMJE) anunció que las revistas médicas importantes no publicarían ensayos clínicos aleatorizados que no hubiesen sido registrados previamente[55].

Detección del sesgo de publicación

Existen formas directas , cuando se detectan estudios que no han sido incluidos en la revisión sistemática y formas indirectas. Entre estas últimas las más empleadas son los gráficos de embudo que ilustran sobre la dispersión de las estimaciones de los efectos de la intervención de los estudios individuales frente a alguna medida del tamaño o la precisión del estudio.

En la actualidad suele emplearse el error estándar de la estimación del efecto de la intervención en el eje vertical [56]. En el horizontal se colocan las estimaciones del efecto. Las medidas de proporción de los efectos de la intervención deben colocarse en una escala logarítmica. Las estimaciones de estudios pequeños se dispersarán en la parte inferior del gráfico. En ausencia de sesgo el gráfico tiene la apariencia de un embudo invertido. Si existe sesgo podría encontrarse una asimetría.

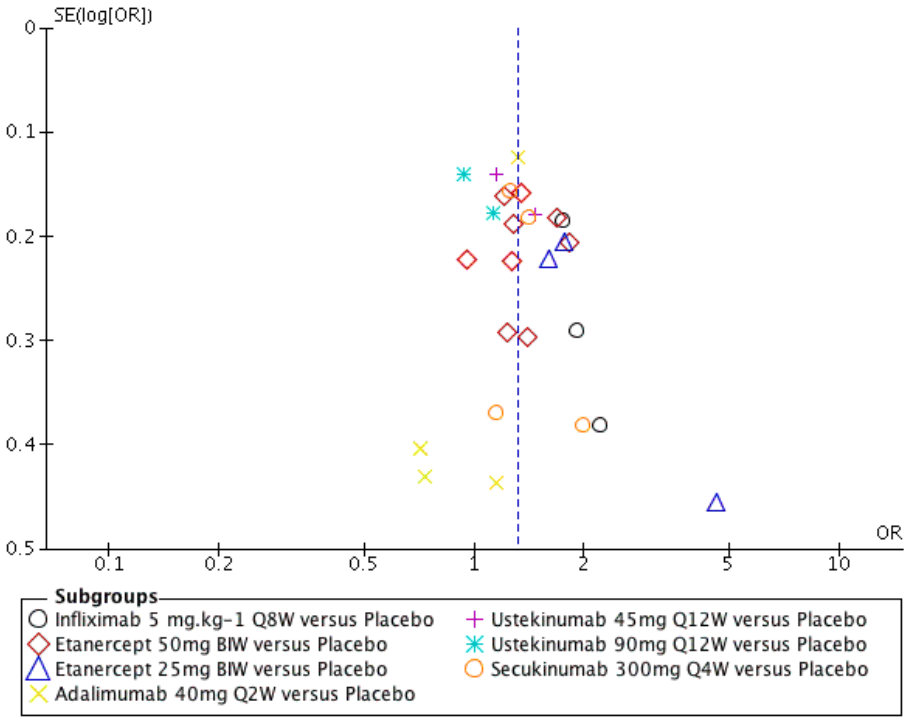


Fig. 7 Ejemplo de funnel plot para evaluar el sesgo de publicación.

Sin embargo este sesgo no siempre provoca asimetría y cuando la hay puede ser explicada por otros motivos como:

- La estimación de los estudios pequeños tiende a diferir de los grandes.
- Los estudios más pequeños tienden a ser de menor calidad metodológica.
- La heterogeneidad verdadera de los efectos de la intervención.
- El azar.

Además, la interpretación visual presenta limitaciones asociadas a la subjetividad[57]. Para paliar este inconveniente, se han desarrollado gráficos de embudo de contorno mejorado que incluyen líneas que se corresponden con los límites de significación estadística y pruebas estadísticas que examinan si la asociación entre la estimación de los efectos de la intervención y la medida del tamaño del estudio es mayor de lo que podría esperarse por el azar. Estas pruebas deben realizarse sólo cuando al menos se hayan incluido 10 estudios en la revisión sistemática y nunca si los estudios incluídos son de tamaño similar. A pesar de todo, no puede excluirse el sesgo cuando no hay evidencia de asimetría dado el bajo poder estadístico de estos test. Entre estas pruebas , la más empleada es la de Egger para resultados continuos con efectos de la intervención medidos como diferencia de medias[60].

Evaluación del grado de calidad de la evidencia

El análisis de la calidad de la evidencia determina el grado de confianza de las estimaciones de los efectos de cada intervención. En la actualidad se aconseja emplear la metodología **GRADE** (*Grades of Recommendation, Assessment, Development, and Evaluation*) y el software **GRADEpro**[10].

El abordaje GRADE establece el grado de calidad para cada variable resultado teniendo en cuenta el conjunto de los estudios primarios considerados para cada una de ellas. Se basa en 3 aspectos:

- **El diseño del estudio.** La confianza en los resultados de los ensayos disminuye si se detectan limitaciones claras en su diseño. De este modo, los ensayos clínicos suelen partir de una calidad moderada o alta y los observacionales de una calidad baja o muy baja.
- **Factores que disminuyen el grado de confianza:**
  - Riesgo de sesgo o error sistemático;

[10] https://gradepro.org/

  - Imprecisión, que implica poca confianza en las estimaciones del efecto y se asocia al número de pacientes y de eventos;
  - Inconsistencia, cuando existen grandes diferencias en la estimación del efecto entre los estudios, se asocia con la heterogeneidad;
  - Problemas de aplicabilidad de la evidencia, cuando no se disponen de comparaciones directas o existen diferencias entre las poblaciones o intervenciones de nuestras preguntas y las disponibles en la literatura relevante;
  - Sesgo de publicación, cuando sospecha de que no se hayan publicado todos los estudios que responden a una pregunta de investigación.

- **Factores que aumentan el grado de confianza** (sólo se puede subir el nivel si no existe ningún factor que disminuya el grado de confianza):
  - Gran magnitud del efecto;
  - Variables de confusión que refuercen la conclusión;
  - Existencia de un gradiente de efecto dosis-respuesta.

La integración de las evaluaciones de los factores anteriores da lugar a cuatro posibles niveles de evidencia para cada variable resultado [59]:

- **Grado de evidencia alto:** Es difícil que los resultados de nuevos estudios modifiquen la confianza en la estimación del efecto.
- **Grado de evidencia moderado:** La confianza en la estimación del efecto y su magnitud podrían cambiar con nuevos estudios.
- **Grado de evidencia bajo:** es probable que nuevos estudios modifiquen la confianza en la estimación del efecto y su magnitud.
- **Grado de evidencia muy bajo:** cualquier estimación del efecto es muy incierta.

La estimación del efecto (su magnitud y dirección) y el grado de calidad de dicha evidencia son el producto final de las revisiones sistemáticas. Las guías de práctica clínica emplean esta información junto con el balance riesgo-beneficio, los valores y preferencias de los pacientes y las consideraciones de los recursos disponibles para establecer la fuerza de las recomendaciones[61].

El sistema GRADE permite sistematizar la toma de decisiones. Cuando la calidad de la evidencia es moderada o alta, se pueden esperar recomendaciones fuertes a favor o en

contra de la intervención; cuando la calidad es baja o muy baja, la mayoría de las veces se puede esperar recomendaciones débiles. Una característica de GRADE es que permite la elaboración de recomendaciones fuertes en contexto de calidad de la evidencia baja o muy baja aunque las circunstacias para ello son excepcionales [62].

II.d. Notificación de los resultados de revisiones sistemáticas y meta-análisis

La claridad, transparencia y consistencia de la presentación de los resultados de las revisiones sistemáticas es fundamental para asegurar su validez.La falta de comunicación de un proceso no implica que no se haya realizado pero pone en riesgo la utilidad de la revisión [63]. En este sentido, se han desarrollado diferentes protocolos que facilitan la tarea de comunicación de las revisiones sistemáticas. El que se desarrolló en primer lugar fue **QUOROM** (*QUality Of Reporting Of Meta-analysis*) en 1999 [64]. Desde entonces la metodología de la notificación ha evolucionado y su denominación ha cambiado a **PRISMA** (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*). El desarrollo de PRISMA comprende desde 2005 hasta su publicación final en 2009[29]. Desde entonces están disponibles seis extensiones de PRISMA que son específicas para diferentes aspectos o tipos de revisiones sistemáticas.

Vamos a repasar los aspectos generales de PRISMA, así como los más específicos de la notificación de los meta-análisis en red (PRISMA-NMA) y de los abstracts de dichos documentos (PRISMA-A).

PRISMA

PRISMA consiste en una lista de comprobación, recogida en la tabla 3, de 27-items divididos en 7 secciones (título, resumen, introducción, métodos, resultados, discusión y financiación).

- **Título (1 item):** El estudio siempre debe identificarse como revisión sistemática o meta-análisis. Esto mejora la calidad de la indexación y de la recuperación del artículo. Existe evidencia de que un 50 % de los casos no se identifica el trabajo correctamente [63]. Se aconseja que refleje los componentes PICO de la pregunta de investigación.

- **Abstract (1 item):** Proporciona información clave de la revisión sistemática. Los abstract estructurados permiten un mejor acceso a la información[65]. Deben resumir todos los aspectos clave de la misma e incluir el registro del protocolo. La principal limitación del abstract es la extensión permitida por las revistas. Existe una extensión de PRISMA desarrollada específicamente para los abstract (PRISMA-A)[66].

- **Introducción (2 items):**

  - Justificación. Debe recoger las razones de su realización, incluyendo la importancia de la pregunta de revisión, el estado actual del conocimiento y limita-

ciones de la evidencia actual. Debe especificarse si es una revisión nueva o una actualización y las razones de ésta.

- Objetivos. Deben indicarse de forma precisa y explícita para informar del alcance y aplicabilidad de la revisión sistemática [67], y reflejar los componentes PICO de la pregunta de investigación.

- **Métodos (12 items):**

  - Protocolo y registro. El objetivo es demostrar que la investigación se ha realizado de forma prospectiva. Debe indicarse si existe, si es posible tener acceso a él y el número de registro. Existen diferentes repositorios donde pueden registrarse protocolos como los de la Universidad de York (PROSPERO) o el del Instituto Joanna Briggs.También pueden publicarse en revistas. En caso contrario debería estar disponible por si es requerido.
  - Criterios de elegibilidad. Deben especificarse inequívocamente y reflejar los componentes PICO de la pregunta . Dado que un estudio se puede describir en varios informes y que un informe puede describir varios estudios es necesario informar tanto de las características del estudio como las del informe. Debe proporcionarse una lista con los estudios excluidos y las causas de exclusión [68].
  - Fuentes de información. Debe informarse: el nombre de las bases de datos empleadas, la plataforma o el proveedor y las fechas de inicio y finalización de la búsqueda de cada una de ellas. También debe notificarse quién desarrolló y llevó a cabo la búsqueda[69].
  - Estrategia de búsqueda. Permite evaluar la exhaustividad de la búsqueda y reproducirla. Se aconseja informar al menos la de una base de datos importante e indicar cómo se tuvieron en cuenta otras. Si se utilizan distintas búsquedas para diferentes partes de una pregunta más amplia se recomienda proporcionar al menos un ejemplo de estrategia para cada parte del objetivo. También se aconseja declarar si las estrategias de búsqueda fueron revisadas por pares.En caso de que las restricciones de espacio no hagan posible publicar todas las estrategias de búsqueda se recomienda: un apéndice o un enlace electrónico a un archivo. Además se aconseja archivarlas para permitir ser replicadas y facilitar actualizaciones futuras.Debe informarse cualquier limitación relacionada con el idioma , fecha o formato de la publicación.

- Selección de los estudios. Se aconseja informar la forma en que se examinan los registros recuperados (normalmente un título y un resumen), la frecuencia con la que fue necesario revisar la publicación de texto completo, si se excluyó cualquier tipo de registro y cómo se llevó a cabo el control del proceso por los autores. PRISMA proporciona un diagrama de flujo para ilustrar el proceso de selección.
- Extracción de los datos. No existe un método estandarizado. Se aconseja describir los métodos empleados (formularios, extracción por duplicado independiente o no, solución de desacuerdos) y las medidas adoptadas para reducir los sesgos y los errores. [70]. Se aconseja informar:
  * si se intentó contactar con los investigadores, lo que pidieron y el éxito en obtener la información necesaria;
  * si se obtuvieron datos de los pacientes individuales de investigaciones originales e indicar los estudios para los que dichos datos se utilizaron en los análisis;
  * confirmar la exactitud de la información incluida. Con respecto a las publicaciones duplicadas se aconseja describir los pasos empleados para evitarlas y reunir los datos de varios informes del mismo estudio.
- Variables. Es importante notificar qué información se buscó, aunque no estuviera disponible [71] y si algunas variables se añadieron después del inicio de la revisión sistemática.
- Riesgo de sesgo de estudios individuales. Deben describirse los métodos utilizados para medir el riesgo de sesgo en los estudios incluidos y cómo se utilizó esa información . Los autores deberían proporcionar una justificación si no se ha realizado una evaluación del riesgo de sesgo.
- Medidas resumen. Deben pre-especificarse: los resultados de interés primario , la medida de efecto resumen para cada resultado y el porqué de la elección de dichas medidas de efecto. Aunque no siempre es fácil juzgar por adelantado que medida es la más adecuada.
- Síntesis de resultados. Se recomienda informar:
  * si se ha realizado transformación de datos;
  * cómo se evaluó la variabilidad entre estudios (heterogeneidad o inconsistencia);

        - cuando se realiza el meta-análisis los autores deben especificar la medida del efecto seleccionada, el método estadístico empleado y las razones de dichas elecciones.

    - Riesgo de sesgo entre los estudios. Se recomienda informar cualquier método utilizado para investigar posibles sesgos entre los estudios. Deben notificarse los resultados de las pruebas de análisis del sesgo de publicación y del sesgo de notificación selectiva.
    - Análisis adicionales. Deben notificarse las pruebas que ayudan a explorar la solidez de los resultados, porqué se hicieron y si estaban preespecificados. Éstas incluyen: análisis de sensibilidad, de subgrupos y meta-regresión [72].

- **Resultados (7 items):**

    - Selección de estudios. Se aconseja informar mediante un diagrama de flujo, el número total de registros identificados de fuentes bibliográficas electrónicas , búsquedas manuales de diversas fuentes, listas de referencias, índices de citas y expertos. El diagrama de flujo y el texto deben describir claramente el proceso de selección y la presencia de informes duplicados o suplementarios.
    - Características de los estudios. Deben recogerse las caracterśticas de los elementos de PICO. De las intervenciones no farmacológicas, puede ser útil especificar los elementos clave de la intervención recibida por cada grupo .Los autores deben proporcionar una cita para la fuente de su información independientemente de si el estudio se publica o no. Las características a nivel de estudio se presentan como una tabla, lo que asegura que se tratan todos los ítems pertinentes y que la información que falta o que no está clara se indica y se acompaña de una revisión narrativa.
    - Riesgo de sesgo entre los estudios. El mejor enfoque es informar explícitamente las características metodológicas evaluadas para cada estudio. Esto puede acompañarse de un texto relevante de los estudios originales que respalde las evaluaciones como se hace en la herramienta de riesgo de sesgo de la Cochrane.
    - Resultados de los estudios individuales. Debe estar justificado que información se va a presentar.Es importante mostrar el efecto estimado con un intervalo de confianza. Esta información debe incluirse incorporada en una tabla o en un gráfico en el que los elementos clave son las estimaciones de efectos , los intervalos de confianza para cada estudio y los datos numéricos del grupo específico. Se debe proporcionar toda la información anterior para cada resultado

incluyendo tanto los beneficios como los daños. Cuando hay muchos resultados deben exponerse los más importantes en el informe principal explicitando si no pueden presentarse debido a la falta de información, con otra información proporcionada como un apéndice Web.

- Síntesis de resultados. Los resultados deben presentarse de manera ordenada. Si los autores han realizado meta-análisis, deben presentarlos como un efecto estimado entre los estudios con un intervalo de confianza. Se recomienda mostrar los resultados reales de los estudios incluidos en una parcela forestal [73] y proporcionar una medida de la consistencia. Las inferencias cualitativas deben presentarse de la forma más sistemática posible con una explicación de por qué no se realizó el meta-análisis.Los autores deben, en general, presentar síntesis de todas las medidas de resultado que se propusieron investigar .Si se aborda una pregunta amplia con un número muy grande de resultados o si los resultados sólo han sido reportados en uno o dos estudios puede ser razonable no presentar todas las medidas de resultado.
- Riesgo de sesgo entre los estudios. Deben presentarse los resultados de la evaluación. Si se informa de un gráfico en embudo, los autores deben especificar la estimación del efecto y la medida de precisión utilizados, presentados típicamente en el eje x e y, respectivamente. También deben describir si han probado la significación estadística de cualquier posible asimetría e informar los resultados de cualquier investigación sobre la notificación selectiva de resultados. Igualmente se debe informar si no se han completado análisis preespecificados para evaluar el riesgo de sesgo entre los estudios y las razones.
- Análisis adicionales. Los autores deben comunicar cualquier análisis de subgrupos o de sensibilidad y si éstos fueron o no preespecificados. Debi'endose informar cualquier prueba de interacciones, así como estimaciones e intervalos de confianza dentro de cada subgrupo. Los resultados de la meta-regresión deben incluir los tamaños de los efectos y los intervalos de confianza [48]. La cantidad de datos incluidos en cada análisis adicional debe especificarse si es diferente de la considerada en los análisis principales.

- **Discusión (3 items):**
  - Resumen de la evidencia. Los autores deben proporcionar un resumen de la naturaleza y conclusiones de la revisión. Los resultados para los cuales se encontraron pocos o ningún dato deben ser anotados debido a la relevancia potencial para las decisiones y la investigación futura.Debe mencionarse la

aplicabilidad de los resultados en diferentes escenarios.Puede presentarse la fuerza de sus recomendaciones resumidas vinculadas a las evaluaciones de la calidad de la evidencia.

- Limitaciones. La discusión de las limitaciones debe abordar: 1) la validez y la presentación de informes de los estudios (riesgo de sesgo, estimaciones del efecto de la intervenciń demasiado imprecisas o falta de datos para participantes o resultados importantes), 2)las limitaciones del proceso de revisión (de la búsqueda , procesos de selección, evaluación y análisis del estudio y 3) la generalización de la revisión(existencia de datos limitados para ciertas poblaciones o subgrupos en los que la intervención podría funcionar de manera diferente)o pocos estudios que evalúen los resultados de interés más importantes; o si hay una cantidad sustancial de datos relativos a una intervención o comparador obsoletos o una fuerte dependencia de la imputación de los valores faltantes para las estimaciones resumidas.
- Conclusiones. Se debe tratar de relacionar los resultados con otras pruebas. Si no se pueden sacar conclusiones debe declararse ya que este hallazgo puede ser tan importante como encontrar efectos consistentes. Debe describirse la información adicional relevante para los tomadores de decisiones, como la rentabilidad de la intervención. Se aconseja hacer recomendaciones explícitas para futuras investigaciones.

- **Financiación (1 item):** Debe exponerse cualquier financiación recibida y el papel de los financiadores, o declarar si la revisión no fue financiada. Además los autores deberían informar cualquier conflicto de interés relacionado con su función o el papel del financiador en la presentación de informes [74].

PRISMA-NMA

Esta extensión de PRISMA fue publicada en 2015 con el objetivo de incorporar los nuevos conceptos y terminologías de las revisiones sistemáticas que sustentan los meta-análisis en red a la disciplina de la síntesis de la evidencia científica[75]). PRISMA-NMA, consta de 32 ítems (5 ítems nuevos y 11 que son modificaciones de los previos de PRISMA).En la tabla 4 se expone con más detalle la lista de verificación de PRISMA para meta-análisis en red.

Entre los conceptos nuevos que incorpora PRISMA-NMA se encuentran:

1. Geometría de la red de tratamientos.

Se refiere a la estructura y forma de la red de tratamientos. Ayuda a establecer la idoneidad de las comparaciones presentes en los estudios y permite detectar si para alguna

de ellas existe escasa o ninguna evidencia. Se representa mediante un grafo que conecta nodos (intervenciones) mediante aristas (estudios). Una vez representado, deben discutirse las características topográficas del mismo (número de estudios, número de pacientes para las comparaciones, etc).

El concepto de geometría de la red da lugar a 3 ítems nuevos que se reparten entre los apartados de metodología y la presentación de resultados:

- Geometría de la red (S1). Aconseja describir los métodos utilizados para explorar la geometría de la red de tratamiento y los posibles sesgos relacionados con ella incluyendo la forma en la que la base de evidencia se ha resumido gráficamente para su presentación.

- Presentación de la estructura de la red (S3). Recomienda proporcionar un gráfico de la red de los estudios incluidos para permitir la visualización de la geometría de la red.

- Resumen de la geometría de la red (S4). Se debería proporcionar una breve descripción de las características de la red de tratamiento. Puede incluir el comentario sobre la cantidad de ensayos y pacientes aleatorizados para las diferentes comparaciones de la red, las lagunas de evidencia y los posibles sesgos reflejados en la estructura de la red.

2. Inconsistencia.

Deriva de los desacuerdos entre los efectos de los tratamientos procedentes de las comparaciones directas e indirectas.

3. Transitividad.

Describe cuándo los diferentes estudios son comparables por no diferir en la distribución de factores modificadores del efecto (diseño del estudio, intervenciones evaluadas, tratamientos concomitantes, la gravedad de los pacientes, etc.). La falta de transitividad podría producir inconsistencias entre la evidencia directa e indirecta y los resultados del meta-análisis en red podrían ser de dudosa utilidad para la toma de decisiones [75].

La notificación de la consistencia y la transitividad da lugar a dos ítems nuevos:

- Evaluación de inconsistencia (S2). En el que se aconseja describir los métodos estadísticos utilizados para evaluar la conformidad de la evidencia directa e indirecta en la/s red/es de tratamiento estudiada/s así como las medidas adoptadas para hacer frente a su presencia cuando se encontró.

- Exploración de la inconsistencia (S5). Describir los resultados de las investigaciones de inconsistencia. Puede incluir información como las medidas de ajuste del modelo

para comparar los modelos de consistencia e inconsistencia, los valores de p de las pruebas estadísticas, o el resumen de las estimaciones de inconsistencia de diferentes partes de la red de tratamiento.

PRISMA-A

La idea de estructurar los abstracts se introdujo hacia finales de los años 90 con la intención de ofrecer una visión más completa y organizada de la información y facilitar el acceso a los hallazgos de las revisiones sistemáticas[76]. Desde ese momento, se ha demostrado que se necesita mejorar la calidad de los resúmenes de las revisiones sistemáticas[77]. La declaración *PRISMA for Abstracts* se publicó en 2013, tratando de ajustarse a cualquier conjunto de encabezamientos de las revistas o la presentación de conferencias [66]. A pesar de que PRISMA-A pone el énfasis en revisiones sistemáticas con meta-análisis sobre estudios de intervención, puede ser empleada en cuestiones de etiología, diagnóstico o test de exactitud de las pruebas. PRISMA-A consta de 12 ítems estructurados en 6 encabezamientos (título, introducción, metodología, resultados, discusión y otros). En la tabla 5 se expone con más detalle la lista de verificación de PRISMA-A.

III. Metaepidemiología y control de calidad de los documentos de síntesis de la evidencia

Introducción.

Tomar decisiones basadas en las mejores pruebas tiene como objetivo optimizar los resultados en salud [78]. Las revisiones sistemáticas son el estándar para resumir de forma científica la mejores pruebas que responden a una pregunta de investigación [8] y sirven de base para el desarrollo de guías de práctica clínica.

Sin embargo, se han encontrado limitaciones metodológicas, tanto en estudios primarios como en las revisiones sistemáticas, que explican diferencias entre los resultados aportados por las pruebas y los obtenidos en la práctica clínica[79]. Del intento de conocer de modo empírico cuáles son dichas limitaciones surge un nuevo área de investigación: la meta-epidemiología. El enfoque de esta nueva disciplina implica el uso de meta-análisis que comparan las estimaciones del efecto de intervención entre ensayos con y sin una característica en particular, como por ejemplo la ocultación de la secuencia de aleatorización [80]. La meta-epidemiología investiga los resultados conflictivos de las revisiones sistemáticas que parten de la misma hipótesis, así como los problemas inherentes al proceso de investigación, como la heterogeneidad, el sesgo de publicación o el ocultamiento de la asignación que dificultan la justificación de los resultados de una revisión sistemática y la elaboración de conclusiones apropiadas, respectivamente [82, 81]. El objetivo final de estos estudios es la translación de sus hallazgos a la investigación primaria y secundaria para mejorar los resultados de las pruebas que dan sustento a las recomendaciones clínicas.

Una de las aplicaciones de la metaepidemiología es la construcción de herramientas de evaluación de la calidad científica tanto de los estudios primarios como secundarios a partir de los conceptos inferidos de la evidencia empírica. Estas herramientas permiten la evaluación de la validez en la revisiones sistemáticas y la detección de errores sistemáticos en las mismas.

La **calidad metodológica** se ha definido como la probabilidad de que el diseño de un estudio genere resultados sesgados [83]. En este sentido, una alta calidad metodológica es el pre-requisito para asegurar la validez de los hallazgos obtenidos en la revisiones sistemáticas [84]. Sin embargo, el cumplimiento de los mejores estándares en la conducción y comunicación de dichos resultados no elimina el riesgo de sesgo que incluye también las desviaciones sistemáticas de la verdad en las estimaciones y/o en las conclusiones de la revisiones sistemáticas [85].

En los últimos años se observa una producción masiva de revisiones sistemáticas y meta-análisis que más que contribuir a la mejora de la evidencia científica son empleadas

como unidades de publicación por su prestigio o como herramientas de marketing [14]. Por ello, poder evaluar la calidad metodológica y el riesgo de sesgo resulta de gran importancia. Una de las aplicaciones de la meta-epidemiología es la construcción de herramientas de evaluación de la calidad científica tanto de los estudios primarios como secundarios a partir de los conceptos inferidos de la evidencia empírica. Estas herramientas permiten la evaluación de la validez y la presencia de errores sistemáticos. Desde hace más de dos décadas se han desarrollado más de 40 herramientas de evaluación de la calidad sin que ninguna esté universalmente aceptada. De todas ellas, la más empleada es **AMSTAR** (*A Measurement Tool to Assess the Methodological Quality of Systematic Reviews*). AMSTAR ha demostrado tener un gran acuerdo interevaluador, ser fiable, factible y válida en la medición de la calidad de las revisiones sistemáticas[86]. Posteriormente, los avances en la evaluación de la calidad de las revisiones sistemáticas y de los estudios primarios han servido de base para el desarrollo en 2016 de ROBIS, la primera herramienta dirigida a analizar el riesgo de sesgo de las revisiones sistemáticas[87].

Hasta el momento no existen estudios que evalúen la calidad metodológica y el riesgo de sesgo de las revisiones sistemáticas publicadas sobre psoriasis. Nuetro trabajo ha empleado AMSTAR y ROBIS para la evaluación de la calidad metodológica y del riesgo de sesgo de este tipo de documentos de síntesis sobre psoriasis en general y sobre intervención, respectivamente. Además se han analizado metadatos que pueden ayudar a predecir la calidad metodológica de las revisiones sistemáticas y meta-análisis sobre psoriasis.

Repasaremos las definiciones y evidencia sobre la meta-epidemiología y posteriormente desarrollaremos los conceptos de evaluación de la calidad metodológica y del riesgo de sesgo de las revisiones sistemáticas, centrándonos en las herramientas AMSTAR y ROBIS.

Meta-epidemiología, Meta-meta-epidemiología y meta-epidemiología en red

El término **meta-epidemiología** fue introducido en 1997 basado en la idea de que la heterogeneidad encontrada en los ensayos clínicos no se debe únicamente al azar, sino que otros factores como diferencias en el diseño, tratamientos, población o los objetivos podían explicarla, al menos en parte[88]. El concepto de meta-epidemiología ha evolucionado con el empleo de métodos estadísticos que examinan la influencia de los problemas cualitativos de los ensayos clínicos[89] hasta el estudio de meta-variables con el fin de controlarlas. Para ello se basa en la metodología de la investigación epidemiológica tradicional siendo los informes de los estudios primarios, en este caso, la unidad de análisis[90]. Por lo tanto, la meta-epidemiología representa la combinación de dos conceptos: epidemiología y meta-análisis.

Los **objetivos** de la meta-epidemiología son[81]:

- Describir la distribución de la evidencia de la investigación para una pregunta específica.
- Examinar la heterogeneidad y los factores de riesgo asociados;
- Controlar el sesgo entre los estudios;
- Resumir la evidencia de la investigación.

No obstante, la meta-epidemiología también tiene una serie de **limitaciones**[91]:

- La meta-epidemiología no puede manejar resultados continuos.
- El poder estadístico de la meta-epidemiología es limitado.
- La meta-epidemiología no puede ser aplicada a las comparaciones indirectas.

Con el objetivo de superar estas limitaciones, se esta desarrollando la **meta-meta-epidemiología** que combina los resultados de varios estudios metaepidemiológicos y la **meta-epidemiología en red**[82]. Estos nuevos conceptos han sido discutidos y pueden diferenciarse entre sí por [91]:

- Las fuentes de obtención de los datos.
- Las restricciones o limitaciones en la selección de las unidades de estudio.
- El tipo de factor de riesgo evaluado.
- La interpretación de la dirección del sesgo y la estimación del impacto en las comparaciones realizadas.
- La asunción respecto a la intercambiabilidad del impacto de los factores de riesgo en las estimaciones del efecto de la intervención.

Evidencias epidemiológicas y desarrollo epistemológico.

El campo de la meta-epidemiología ha experimentado un gran crecimiento en los últimos 5 años en los que se ha publicado el 60% de lo estudios sobre este nuevo área de investigación[92]. Los primeros trabajos meta-epidemiológicos se centraron en el control de la influencia de la ocultación de la asignación y el cegamiento post-aleatorización en los ensayos clínicos [90]. Se demostró que las evaluaciones subjetivas de sus efectos fueron exageradas.

A partir de 2008 se amplía la aplicación de las técnicas metaepidemiológicas al estudio de otras variables de metaconfusión como:

- El genotipo[93];
- El diseño del estudio[94];
- El número de participantes [95];
- La diferencia entre el tamaño del efecto en ensayos clínicos uni o multicéntricos;
- Las características de la publicaciones (tipo de fuente, lenguaje, indexación en bases de datos electrónicas);
- Las características de los ensayos clínicos (aleatorización desigual, cruzados, paralelos, cese precoz del ensayo, ensayos con potencia no informada)
- Las características referentes a la población(simples vs multicéntricos, adultos vs niños o ancianos);
- La publicación (conflictos de intereses, tipo de financiación)[92].

Hasta el momento, se ha observado que una gran heterogeneidad en la conducción y publicación de las revisiones sistemáticas puede resumirse en los siguientes puntos: 1)La selección de revisiones sistemáticas y la extracción de los datos. 2) Las técnicas de síntesis y análisis. 3) Las medidas empleadas para controlar la confusión.

Por todo ello es necesario la estandarización epistemológica. Si bien la mayoría de los trabajos meta-epidemiológicos de alta calidad han adoptado la metodología de las revisiones sistemáticas, sería deseable elaborar un manual para su conducción y protocolos de notificación para este tipo de estudios, de forma similar por ejemplo al Manual Cochrane [8] y a la declaración PRISMA [29]. Esto mejoraría la validez de los trabajos meta-epidemiológicos.

Además, las evaluaciones comparativas de las diversas alternativas para cada paso meta-analítico ayudarían a identificar qué métodos son más robustos y preferibles. En la actualidad, son escasas las investigaciones en este sentido y mayoritariamente están basadas en estudios de simulación más que en bases de datos empíricas reales. En este sentido se están produciendo avances en este campo:

- El cálculo del tamaño muestral para este tipo de estudios que podría utilizarse para controlar la impresición [96].
- La sistematización de uso de entornos de lenguaje de análisis de datos y creación de gráficos con el lenguaje R[97].
- La investigación de las herramientas de conducción [98].

- Creación de modelos de análisis sensitivo de meta-sesgos como el *modelo paramétrico de Copas* [99].

- La presentación de gráficos[100].

Finalmente sería de gran interés incorporar un enfoque similar a GRADE para analizar la utilidad de los resultados con la calidad de la evidencia[101]. El grado de evidencia meta-epidemiológica puede emplearse para guiar la conducción de las revisiones sistemáticas o incorporarse directamente en herramientas que evalúan la validez interna o el riesgo de sesgo de los ensayos clínicos incluidos en una revisión sistemática [102]. El desarrollo de la herramienta de evaluación de riesgo de sesgo de la Cochrane es un ejemplo en este sentido [44]. A continuación repasaremos la aplicación de dichos conocimientos en el desarrollo de herramientas dedicadas a la evaluación de la calidad metodológica y del riesgo de sesgo de las revisiones sistemáticas, centrándonos en AMSTAR y ROBIS.

Evaluación de la calidad metodológica de las revisiones sistemáticas.

En las dos últimas décadas se ha producido el desarrollo de de múltiples herramientas cuya estructura y rigor varían en función de las siguientes características:

- **Evaluación de la validez:** De las más de 40 herramientas para la evaluación de la calidad metodológica sólamente tres, The Assessment of Multiple Systematic Reviews (AMSTAR), OQAC y la herramienta para la evaluación de la calidad de los meta-análisis han presentado un desarrollo riguroso[85].

- **Tipo de estudios que evalúan:** La mayoría de las herramientas no son específicas para el tipo de revisión sistemática o meta-análisis para el que han sido desarrolladas[85].

- **Estructura de la herramienta:** Varía desde listas de verificación y escalas, que están desaconsejadas en la actualidad, hasta herramientas de evaluación por dominios.

- **Estructura de los ítems/número de ítems:** Varían en complejidad desde la descripción narrativa , sistemas semi-estructurados de puntuación, puntuaciones simples hasta sistemas de más complejos que emplean sumas de escores para un resultado total.

- **Fiabilidad inter-evaluador:** Únicamente 5 herramientas presentaban datos del acuerdo interobservador.

- **Aspectos de las revisiones sistemáticas evaluados:** La mayoría de las herramientas desarrolladas presentan ítems que evalúan cada uno de los aspectos de las revisiones sistemáticas: criterios de selección, búsqueda, revisión, síntesis y conclusiones.

AMSTAR (*A Measurement Tool to Assess the Methodological Quality of Systematic Reviews*)

AMSTAR es una herramienta para evaluar la calidad metodológica de las revisiones sistemáticas, se publicó en 2007 y su desarrollo se realizó sobre la base de:

- dos herramientas previas de evaluación de la calidad: el cuestionario de evaluación de la calidad (OQAQ) y una lista de verificación creada por Sacks;
- la opinión de expertos;
- conceptos derivados de avances metodológicos relacionados con las restricciones del lenguaje , el sesgo de publicación y la inclusión de literatura gris en la revisión sistemática[54].

AMSTAR ha demostrado presentar buen acuerdo, confiabilidad, validez de construcción y factibilidad en la evaluación de revisiones sistemáticas sobre estudios de intervención, aunque posteriormente se ha empleado para otros tipos revisiones. También ha demostrado presentar una buena validez externa [103].

La herramienta, consta de 11 ítems que repasan los aspectos más importantes de la conducción y comunicación de las revisiones sistemáticas. Cada pregunta tiene cuatro posibles respuestas (sí, no, no puede contestarse, no aplicable). La respuesta "sí" otorga un punto y el resto 0 puntos. AMSTAR se puede puntuar individualmente (componentes) o como una lista de verificación sumando las puntuaciones de los ítems (puntuación general). Todas las revisiones sistemáticas tienen la misma probabilidad de puntuar bien, pero las que incorporen meta-análisis tendrán una puntuación superior en el conjunto de resultados. En la tabla 7 se exponen las preguntas, justificación y los razonamientos de apoyo para la toma de decisiones.

Limitaciones de AMSTAR para evaluar la calidad metodológica de las revisiones sistemáticas

- Ausencia de valoración de algunos aspectos importantes de las revisiones sistemáticas.
  - Evaluación de la calidad de la evidencia para cada resultado importante, la confianza en las estimaciones de efecto [104].

  - Realización de análisis de subgrupos de sensibilidad dado que los efectos del tratamiento pueden diferir entre las poblaciones o por las características de las intervenciones [105].

- Opciones de respuesta problemáticas [106]. La expresión "no puede responder" puede ser difícil de interpretar y distinguir de "no" cuando no se proporciona información adicional. Dado que un enfoque común es asumir que si los autores no informaron de un paso, entonces no ocurrió, "no" sería la respuesta apropiada. Además, "no aplicable" sólo es apropiado para los item 9 y 10 cuando estos elementos no son posibles o apropiados; todos los demás elementos deberían ser siempre abordados.

- Problemas con el cálculo de la puntuación total:

  - La orientación para puntuar artículos individuales y obtener una puntuación total es deficiente. Las revisiones sistemáticas cumplen parcialmente los criterios de la búsqueda, pero por cuestiones de espacio de la revista no proporcionan las estrategias de búsqueda o palabras clave. Existen modificaciones de AMSTAR que permiten puntuar los elementos parcialmente cumplidos, como R-AMSTAR [107]. Si bien una revisión sistemática encontró que AMSTAR tenía mejores propiedades de medición de la calidad metodológica que R-AMSTAR[108].
  - AMSTAR no proporciona ninguna guía sobre cómo combinar las puntuaciones individuales de evaluadores múltiples. En este sentido se han promediado las puntuaciones entre los evaluadores para abarcar cada evaluación independiente [109] o se han promediado las puntuaciones entre dos evaluadores cuando existen diferencias de uno o dos puntos y la participación de un tercer evaluador cuando las puntuaciones diferían en tres o más puntos [110].
  - AMSTAR no orienta sobre cómo traducir la puntuación total en calificaciones categóricas [111]. En este sentido, se han empleado varios umbrales, por lo que es difícil comparar las evaluaciones a través de revisiones.
  - En AMSTAR cada ítem tiene el mismo peso a pesar de que no existe evidencia empírica que demuestre que esto es así realmente[84].

- La equivalencia de "no aplicable", "no" y "no se puede contestar" ,todos calificados como cero, es problemática dado que "no aplicable" no debería tenerse en cuenta en la puntuación total[109].

ROBIS (*Risk Of Bias of Systematic reviews*)

ROBIS es la primera herramienta dirigida a evaluar el riesgo de sesgo de las revisiones sistemáticas. Entendiendo ese riesgo como el de cometer un error sistemático o una desviación de la verdad, en las estimaciones resumidas y/o en las conclusiones de la revisiones sistemáticas. Sólo se relaciona con la validez interna de la revisión sistemática y, por lo tanto, no considera aplicabilidad de la misma[85].

ROBIS se estructura en dominios, con un enfoque en tres etapas:

- información, utilizada para apoyar el juicio del riesgo de sesgo;
- preguntas de señalización, que se responden como 'sí', 'probablemente sí', 'probablemente no', 'no' y 'no hay información'.
- juicio final que se evalúa como riesgo de sesgo 'bajo', 'alto' o 'poco claro'.

El desarrollo de la herramienta se orientó en cuatro etapas:

- Definición del alcance. :
  - El grado de coincidencia entre la pregunta de investigación de la revisión sistemática y la que está siendo dirigida por el revisor.
  - El grado en que los métodos de revisión sistemática minimizan el riesgo de sesgo en las estimaciones resumidas y en las conclusiones de la revisión. En este sentido la herramienta distingue entre revisión sistemática de riesgo de sesgo alto, bajo e incierto.
- Revisión de la evidencia. Se utilizaron 3 enfoques:
  - Clasificación de los estándares del protocolo de MECIR.
  - Revisión de las herramientas diseñadas para evaluar la calidad de revisiones sistemáticas o meta-análisis.
  - Revisión de las reseñas generales de AMSTAR para evaluar la calidad de las revisiones sistemáticas con el fin de proporcionar información sobre los potenciales usuarios de ROBIS.
- Reunión de expertos. Consistió en una reunión presencial más un proceso Delphi donde se evaluaron la estructura , dominios y preguntas de señalización.

- Experimentación con la herramienta: Mediante un proceso Delphi modificado se observó:
  - un mejor acuerdo para las revisiones sistemáticas que tienen bajo riesgo de sesgo;
  - las dificultades en la aplicación de ROBIS están relacionadas principalmente con las limitaciones en la presentación de informes de las revisiones sistemáticas más que con las de la aplicación de la herramienta en sí.

En conclusión, el desarrollo de guías de práctica clínica basadas en la evidencia depende de la producción de pruebas válidas. La validez representa la confianza con la que se responde a la pregunta de investigación , validez interna, y la capacidad de extrapolar los resultados de la investigación , validez externa. La meta-epidemiología tiene como objetivos fundamentales:1) ayudar a mejorar la metodología de la conducción de estudios científicos primarios y secundarios, constantemente se producen cambios en las guías en este sentido;2) perfeccionar los puntos clave de la notificación, lo que justifica PRISMA y sus extensiones y 3) evaluar la calidad de las pruebas que sirven de base para realizar las recomendaciones. En este sentido se han desarrollado múltiples herramientas tanto para estudios primarios , Vg, la herramienta de medición del riesgo de sesgo de la Cochrane [44] como secundarios, Vg; herramientas de medición de calidad de las revisiones sistemáticas, AMSTAR [84], o de la evaluación del riesgo de sesgo, ROBIS Tool[87]. Sin embargo la meta-epidemiología al tratarse de una ciencia reciente necesita del desarrollo de su cuerpo epistemológico para unificar los criterios y mejorar la aplicabilidad de sus resultados.

IV. Guías de práctica clínica: recomendaciones basadas en la evidencia para la toma de decisiones.

Tomar decisiones en salud es un proceso complejo. Los clínicos se enfrentan a diario a escenarios en los que disponen de dos o más alternativas que se diferencian en eficacia, seguridad y costes entre otros criterios, sin el tiempo ni los recursos necesarios para la considerar la evidencia subyacente. En este contexto las decisiones pueden basarse en la experiencia personal, en el consejo de compañeros o expertos, o en las recomendaciones de las guías de práctica clínica. Éstas pueden mejorar la calidad de las decisiones, el uso de los recursos y de forma global los resultados en salud [112]. Para ello, deben considerar todos los factores relevantes que influyen en una decisión de forma estructurada, explícita y transparente, y proporcionar recomendaciones claras y realizables [113]. Cuando esto no es así, tienen el riesgo de inducir intervenciones subóptimas, ineficaces o perjudiciales o costosas e inasequibles para los servicios de salud que disminuyen los recursos necesarios para actuaciones más efectivas. En resumen, las guías de práctica clínica traducen la calidad de la evidencia disponible, confianza en los efectos de una intervención, en recomendaciones cuya fuerza depende de la integración de varios factores.

El método y los juicios que deben ser considerados en su desarrollo varían en función del tipo de decisión. Sin embargo, algunos criterios son relevantes para todas, como los efectos y la calidad de la evidencia de las opciones consideradas, las consideraciones económicas o su factibilidad.

Se deben hacer juicios sobre cada opción que estén informados por las mejores pruebas disponibles. En este sentido, se han publicado diferentes sistemas de categorización de la calidad de la evidencia que también realizan recomendaciones como: el modelo inglés del Oxford Centre for Evidence-Based Medicine(OCEBM) [114], el escocés del Scottish Intercollegiate Guidelines Network(SIGN) [115] o el estadounidense del American College of Chest Physicians (ACCP) [116]. En ellos, los niveles de calidad de los estudios permiten establecer distintos grados de recomendación. Sin embargo, estos sistemas se desarrollaron mediante el consenso de opinión de expertos y no han sido validados [117]. Debido a ello se han observado diferencias en la categorización de los niveles de evidencia de los estudios y de los grados de recomendación, siendo algunos mejores en la estimación del nivel de evidencia que en establecer el grado de recomendación, y viceversa. Todo ello representa un riesgo para la fiabilidad de las guías de práctica clínica.

En 2004 se publicó la propuesta del grupo de trabajo GRADE, formado por un conjunto internacional y multidisciplinar de metodólogos, expertos en este tipo de documentos y médicos clínicos, con la intención de disminuir las incertidumbres descritas [118]. GRADE

propone un abordaje diferente basado en los sistemas previos, pero con una mejor estructura y transparencia en la clasificación del nivel de evidencia y el establecimiento de la fuerza de las recomendaciones [119]. El sistema GRADE proporciona tablas de los perfiles de evidencia y sumario de hallazgos [120]. Su uso presenta las siguientes ventajas:

- Separar la calidad de la evidencia de la fuerza de las recomendaciones.
- Valorar la importancia relativa de las variables de resultado o desenlace.
- Proporcionar descripciones detalladas de los criterios de calidad de la evidencia respecto a resultados o desenlaces concretos.
- Emplear definiciones explícitas y juicios secuenciales durante el proceso de categorización.
- Considerar el balance entre beneficios y riesgos, los valores del paciente y el consumo de recursos o costes.

Fuerza de las recomendaciones según GRADE. El sistema GRADE establece la fuerza de las recomendaciones basándose en distintos factores[59]:

- **Balance entre riesgos y beneficios.** La certidumbre o incertidumbre del balance riesgo/beneficio determinará de forma importante la fuerza de la recomendación.
- **Calidad de la evidencia** para cada uno de los desenlaces de interés: los niveles de evidencia son cuatro [59]:
  - *Alto:* Es difícil que los resultados de nuevos estudios modifiquen la confianza en la estimación del efecto.
  - *Moderado:* La confianza en la estimación del efecto y su magnitud podrían cambiar con nuevos estudios.
  - *Bajo:* es probable que nuevos estudios modifiquen la confianza en la estimación del efecto y su magnitud.
  - *Muy bajo:* cualquier estimación del efecto es muy incierta.
- **Los valores y preferencias de los pacientes.** Fortalecen el grado de la recomendación cuando la concordancia es alta y lo debilitan cuando existe una elevada variabilidad .Es necesario realizar un juicio de valor y conocer previamente los valores y preferencias de la población de nuestro entorno y de posibles diferencias individuales. En este sentido, todavía faltan estudios apropiados que los analicen en diferentes situaciones.

- **Estimación del consumo de recursos o costes.** El análisis de costes suele requerir el concurso de expertos en economía de la salud. En general, se considera que una intervención se puede calificar de muy coste-efectiva si cuesta menos de una vez la media de la renta per cápita de un país o región por año de vida ajustado por calidad (AVAC) ganado. Incluso hasta 3 veces la media de la renta per cápita por AVAC ganado puede ser tolerable. La OMS ha desarrollado tablas de umbrales al respecto.[121]

Finalmente, las recomendaciones se gradúan de forma binaria en fuertes (grado 1) o débiles (grado 2), bien a favor o en contra. Las recomendaciones fuertes conllevan el mensaje de que la intervención debería ser ofrecida a todos o casi todos los pacientes si es a favor o que no debería ser usada en ninguno o casi ninguno de los pacientes si es en contra. Por el contrario, una recomendación débil conlleva el mensaje de que lo que se está proponiendo debe ser considerado a la luz de las circunstancias clínicas y las preferencias de los pacientes.

Estructuración de los marcos de decisión de la evidencia (EMDE). Los marcos de referencia para las recomendaciones de práctica clínica proporcionan un enfoque estructurado y transparente, ayudan a considerar los criterios claves que determinan si una intervención debe ser recomendada y que los juicios están apoyados por la mejor evidencia disponible. La estructuración de los marcos de la evidencia se basa en los siguientes aspectos:

- **Formulación de la pregunta.** Cuando se formulan las preguntas debe especificarse: los pacientes, la intervención, la comparación y los resultados (PICO), la perspectiva, los subgrupos para los cuales la evidencia y sus juicios y recomendaciones pueden diferir de una recomendación general y el escenario para la cual se pretende la recomendación [122]. La perspectiva de la pregunta debe ser explicitada porque puede dar lugar a la formulación de distintas recomendaciones. Esto puede conducir a la confusión y/o a recomendaciones inadecuadas. Por ejemplo en un paciente individual, los costos de bolsillo pueden ser críticos para tomar una decisión. Sin embargo, desde una perspectiva poblacional las decisiones afectan la manera en que se utilizan recursos sanitarios limitados. En este sentido, los requerimientos totales de recursos , la rentabilidad desde la perspectiva de la población y los impactos sobre la equidad resultan críticos. Es por ello que el mandato de la organización que formula una recomendación suele determinar la perspectiva específica que adopta un grupo especial: Si se desarrollan directrices nacionales, podría adoptarse la perspectiva del gobierno o el departamento de salud para asegurar el uso óptimo del presupuesto . Sin embargo, una sociedad profesional podría adoptar una perspectiva individual del

paciente .Las recomendaciones pueden diferir entre los subgrupos de la población considerados al formular la pregunta. Esto puede deberse a diferencias en las personas , en las intervenciones, en las comparaciones, o en otros aspectos[123].

- **Evaluación de los criterios considerados.** Los criterios que sirven de base para los marcos de decisión difieren en algunos aspectos para las recomendaciones poblaciones frente a las de pacientes individuales. La evidencia para informar los juicios puede provenir de diferentes fuentes que deben ser anotadas junto con sus limitaciones.
- **Consideraciones adicionales para cada criterio.** Los criterios a tener en cuenta en la toma de decisiones son:
  - *¿El problema es una prioridad?.* Cuanto más grave sea el problema, más probable es que una intervención sea una prioridad o se deba recomendar. Desde el punto de vista de la población, las intervenciones útiles para las afecciones mortales o incapacitantes son probablemente una prioridad más alta y recomendable que las que se producen en condiciones transitorias o las que causan angustia menor y reversible. No obstante, se puede decidir que todos los problemas que aborda una directriz particular son igualmente importantes, haciendo que este criterio sea irrelevante. Desde una perspectiva individual del paciente, la importancia del problema no es relevante , dado que siempre será una prioridad. Sin embargo, la importancia de un problema puede afectar las decisiones tomadas por los pacientes individuales. Por ejemplo, las prioridades de los pacientes para la prevención primaria pueden afectar la fuerza de las recomendaciones, ya que algunos riesgos podrían ser más importantes que otros, o el riesgo basal de los pacientes podría ser tan bajo que la prevención no sería una prioridad.
  - *¿Cuál es la magnitud de los efectos esperados deseables e indeseables?.* Los resúmenes de los hallazgos proporcionan estimaciones de los efectos de las intervenciones que se comparan en los resultados de interés. Los juicios sobre la importancia de los efectos deben tener en cuenta la magnitud absoluta del efecto y la importancia del resultado , cuánto es valorado por las personas afectadas. Cuanto mayores sean los efectos deseables, más probable es que se recomiende una intervención. Y cuanto mayores sean los efectos indeseables menor será la probabilidad de que se recomiende.
  - *¿Cuál es la calidad de la evidencia de los efectos?.*A menor calidad de la evidencia de los principales resultados, menor probabilidad de que se haga una fuerte recomendación para una intervención y más probable es que la intervención deba ser evaluada para su implementación [124].

- *¿Hay incertidumbre o variabilidad importante en la valoración de los resultados principales?.* La incertidumbre sobre la valoración de los resultados de interés por los grupos afectados y la variabilidad e la manera en que los pacientes valoran los principales resultados pueden razones para hacer una recomendaciones débiles.
- *¿El equilibrio entre los efectos deseables e indeseables favorece la intervención o la comparación?.* Los juicios sobre el equilibrio entre los efectos deseables e indeseables deben tener en cuenta juicios anteriores.
- *¿Cuáles son los de recursos necesarios?.* Los costes dependen de la perspectiva que se tome. Cuanto mayor sea el coste, menor será la probabilidad de que se recomiende una intervención. Si se considera que el uso de los recursos es crítico para una decisión sobre una recomendación, es más probable que se evalúe formalmente el uso de dichos recursos.
- *¿Cuál es la calidad de la evidencia para los recursos requeridos?.* Mientras menos segura es la evidencia de los requerimientos de recursos, menos probable es que se haga una recomendación fuerte para o contra una intervención. Las sentencias sobre la certeza de la evidencia de las necesidades de recursos son similares a los juicios sobre la evidencia de los efectos [125].
- *¿La rentabilidad de la intervención favorece a la intervención o la a comparación?.* Cuanto mayor sea el costo en relación con el beneficio neto, menor será la probabilidad de que se recomiende una intervención. Los juicios sobre la rentabilidad de una intervención deben tener en cuenta el equilibrio entre los efectos deseables e indeseables, la certeza de la evidencia de los efectos y la incertidumbre o variabilidad en cuánto se valoran los principales resultados . Además, si se utiliza una relación costo-efectividad de una evaluación económica formal también se debe considerar cómo de fuerte es la estimación con variables únicas o múltiples en el modelo, si la evaluación económica es fiable y si se utiliza una evaluación económica publicada.
- *¿Cuál es el impacto en las acciones de salud?.* Si la intervención puede reducir las desigualdades para las personas o población y existe evidencia de ello mayor será la probabilidad de tener una recomendación fuerte.
- *¿Es aceptable la intervención para las partes interesadas?.* Una intervención puede ser inaceptable en función de la relación de los efectos y costos deseables e indeseables o por el desacuerdo sobre los principios éticos (como la autonomía, la no maleficencia, la beneficencia o la justicia [126]. Cuanto menos aceptable sea una intervención, menos probable es que se recomiende, o si se recomienda,

más probable es que se necesite una estrategia de implementación para abordar las preocupaciones sobre capacidad de aceptación.

– *¿Es viable la implementación de la intervención?.* Cuanto menos factible es una intervención , menos probable es que se deba recomendar. Las dificultades para llevar a cabo una intervención también pueden modificar su fuerza de recomendación. Las guías de práctica clínica pueden ayudar a los responsables de implementar las recomendaciones, abordando los obstáculos claves para su recomendación en sus conclusiones[127].

Integración de las recomendaciones. La evaluación global de los puntos anteriores debe hacer llegar a una conclusión sobre la dirección y la solidez de la recomendación[128]. Las recomendaciones deben justificarse basándose en los criterios empleados en la evaluación. Una justificación resume los juicios para cada uno de los criterios que fueron más importantes para la decisión. Deben especificarse las preocupaciones clave sobre la viabilidad y aceptabilidad de la intervención y las estrategias para abordar esas preocupaciones, los indicadores que deben ser controlados y las prioridades para futuras investigaciones[129]. Puede emplearse el uso de procesos de consenso formales o informales o el voto de personas que intervienen para establecer las recomendaciones en función de la complejidad. Para recomendaciones sencillas los procesos de consenso informales suelen ser suficientes.

La preocupación más importante sobre el uso de marcos de decisión basados en la evidencia es que son complejos y requieren recursos adicionales para la preparación. El reto es mantener el enfoque para hacer estos juicios lo más simple posible. Algunas limitaciones de este tipo de estrategia para la elaboración de guías de práctica clínica son:

- El método se ha desarrollado para responder a cuestiones sobre intervenciones alternativas, sobre todo de tratamiento o prevención, no sobre riesgo o pronóstico, y tiene dificultades respecto a pruebas diagnósticas, temas de salud pública o sistemas de salud.

- El método no considera todos los pasos de las guías de práctica clínica.

- Este sistema no elimina por completo los desacuerdos que pueden existir al valorar una evidencia o al decidir opciones alternativas, puesto que siempre hay una influencia subjetiva en todo juicio.

Hipótesis

Identificar las evidencias obtenidas a través de diferentes ensayos clínicos sobre la eficacia y la seguridad a corto plazo de los fármacos biológicos aprobados para el tratamiento de la psoriasis moderada-severa en adultos podría ser importante para apoyar a los clínicos en la toma de decisiones. El empleo de una revisión sistemática y un meta-análisis en red supondría la mejor estrategia para comparar directa e indirectamente la magnitud y dirección de los efectos estimados para todas las opciones estudiadas, permitiendo realizar comparaciones entre fármacos que no han sido evaluados directamente en ningún ensayo clínico y sugerir un orden de prioridad de uso de dichos fármacos basado en el cálculo de la probabilidad de que sean efices o de que asocien un evento adverso.

El análisis de la calidad metodológica y del riesgo de sesgo de la revisiones sistemáticas y meta-análisis publicados sobre la psoriasis, permitiría identificar qué factores caracterizan aquellos documentos de síntesis de mejor calidad científica y servirían de base para el desarrollo de modelos predictivos útiles tanto para productores como para consumidores de este tipo de documentos.

Objetivos

OBJETIVOS PRIMARIOS:

1. Comparar la eficacia y seguridad a corto plazo de los fármacos biológicos aprobados para el tratamiento de la psoriasis en placas moderada-severa en adultos mediante un meta-análisis en red. Abordado en: Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

2. Analizar la calidad metodológica de las revisiones sistemáticas y los meta-análisis publicados sobre psoriasis.Abordado en: Gómez-García F*, Ruano J*, Aguilar-Luque M, Gay-Mimbrera J, Maestre-Lopez B, Sanz-Cabanillas JL, Carmona-Fernández PJ, González-Padilla M, Vélez García-Nieto A, Isla-Tejera B. **Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest and bibliometric indices as predictors of methodological quality.** *Br J Dermatol.* 2017 Jun;176(6):1633-1644. doi: 10.1111/bjd.15380. Epub 2017 May 19.

3. Analizar el riesgo de sesgo de las revisiones sistemáticas y los meta-análisis publicados sobre psoriasis.Abordado en:Gómez-García F*, Ruano J*, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, Maestre-López B, Carmona-Fernández PJ, González-Padilla M, García-Nieto AV, Isla-Tejera B. **Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool.** *J Clin Epidemiol.* 2017 Sep 9. pii: S0895-4356(17)30534-6. doi: 10.1016/j.jclinepi.2017.08.015. [Epub ahead of print]

OBJETIVOS SECUNDARIOS:

1. Establecer un ranking de eficacia de los fármacos biológicos aprobados para el tratamiento de la psoriasis en placas moderada-severa en adultos basado en la probabilidad de alcanzar PASI 70 o PASI 90 a las 10-16 semanas. Abordado en:Gómez-García F*,

Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

2. Establecer un ranking de seguridad de los fármacos biológicos aprobados para el tratamiento de la psoriasis en placas moderada-severa en adultos basado en la probabilidad de tener un efecto adverso a las 10-16 semanas.Abordado en:Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

3. Analizar con la herramienta Cochrane el riesgo de sesgo de los ensayos clínicos incluidos en la revisión sistemática.Abordado en: Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

4. Analizar con GRADE la calidad de la evidencia de las principales variables analizadas relacionadas con la eficacia y la seguridad.Abordado en:Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

5. Identificar las variables predictoras de alta calidad metodológica de una revisión sistemática sobre psoriasis.Abordado en:Gómez-García F*, Ruano J*, Aguilar-Luque M, Gay-Mimbrera J, Maestre-Lopez B, Sanz-Cabanillas JL, Carmona-Fernández PJ, González-Padilla M, Vélez García-Nieto A, Isla-Tejera B. **Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest and bibliometric indices as predictors of methodological quality.** *Br J Dermatol.* 2017 Jun;176(6):1633-1644. doi: 10.1111/bjd.15380. Epub 2017 May 19.

6. Crear un modelo predictivo de alta calidad metodológica de una revisión sistemática sobre psoriasis. Abordado en:Gómez-García F*, Ruano J*, Aguilar-Luque M, Gay-Mimbrera J, Maestre-Lopez B, Sanz-Cabanillas JL, Carmona-Fernández PJ, González-Padilla M, Vélez García-Nieto A, Isla-Tejera B. **Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest and bibliometric indices as predictors of methodological quality.** *Br J Dermatol.* 2017 Jun;176(6):1633-1644. doi: 10.1111/bjd.15380. Epub 2017 May 19.

7. Identificar los ítems de AMSTAR que permiten discriminar mejor entre revisiones sistemáticas de alta y baja calidad metodológica.Abordado en:Gómez-García F*, Ruano J*, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, Maestre-López B, Carmona-Fernández PJ, González-Padilla M, García-Nieto AV, Isla-Tejera B. **Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool.** *J Clin Epidemiol.* 2017 Sep 9. pii: S0895-4356(17)30534-6. doi: 10.1016/j.jclinepi.2017.08.015. [Epub ahead of print]

8. Identificar las preguntas de señalización de ROBIS que permiten discriminar mejor entre revisiones sistemáticas de bajo y alto riesgo de sesgo.Abordado en:Gómez-García F*, Ruano J*, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, Maestre-López B, Carmona-Fernández PJ, González-Padilla M, García-Nieto AV, Isla-Tejera B. **Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool.** *J Clin Epidemiol.* 2017 Sep 9. pii: S0895-4356(17)30534-6. doi: 10.1016/j.jclinepi.2017.08.015. [Epub ahead of print]

9. Analizar el grado de correspondencia entre los ítems de AMSTAR y ROBIS al evaluar la calidad metodológica y el riesgo de sesgo de las revisiones sistemáticas sobre psoriasis.Abordado en:Gómez-García F*, Ruano J*, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, Maestre-López B, Carmona-Fernández PJ, González-Padilla M, García-Nieto AV, Isla-Tejera B. **Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool.** *J Clin Epidemiol.* 2017 Sep 9. pii: S0895-4356(17)30534-6. doi: 10.1016/j.jclinepi.2017.08.015. [Epub ahead of print]

Artículos

Durante mi periodo doctoral llevé a cabo varios análisis, cuyos resultados dieron lugar a un número de artículos y manuscritos. Tres de esos artículos han sido seleccionados para la evaluación de mi proyecto de tesis:

- Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.
  Índices bibliométricos: *Journal Impact factor (2016):* 4.706. *Rank (Dermatology, 2016):* 5/63 (Q1/D1).

- Gómez-García F*, Ruano J*, Aguilar-Luque M, Gay-Mimbrera J, Maestre-Lopez B, Sanz-Cabanillas JL, Carmona-Fernández PJ, González-Padilla M, Vélez García-Nieto A, Isla-Tejera B. **Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest and bibliometric indices as predictors of methodological quality.** *Br J Dermatol.* 2017 Jun;176(6):1633-1644. doi: 10.1111/bjd.15380. Epub 2017 May 19.
  Índices bibliométricos: *Journal Impact factor (2016):* 4.706. *Rank (Dermatology, 2016):* 5/63 (Q1/D1).

- Gómez-García F*, Ruano J*, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, Maestre-López B, Carmona-Fernández PJ, González-Padilla M, García-Nieto AV, Isla-Tejera B. **Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool.** *J Clin Epidemiol.* 2017 Sep 9. pii: S0895-4356(17)30534-6. doi: 10.1016/j.jclinepi.2017.08.015. [Epub ahead of print]
  Índices bibliométricos: *Journal Impact factor (2016):* 4.978. *Rank (Health Care Sciences and Services, 2016):* 6/90 (Q1/D1).

Otros artículos sobre el proyecto, de los que soy coautor y que no han sido incluídos finalmente en este trabajo:

- Juan Luis Sanz-Cabanillas[*], Juan Ruano[*], Francisco Gomez-Garcia, Patricia Alcalde-Mellado, Jesus Gay-Mimbrera, Macarena Aguilar-Luque, Beatriz Maestre-Lopez, Marcelino Gonzalez-Padilla, Pedro J. Carmona-Fernandez, Antonio Vélez García-Nieto, Beatriz Isla-Tejera. **Author-paper affiliation network architecture influences the methodological quality of systematic reviews and meta-analyses of psoriasis.** *PLoS One*. 2017 Apr 12;12(4):e0175419. doi: 10.1371/journal.pone.0175419. eCollection 2017. [Epub ahead of print]
  Índices bibliométricos: *Journal Impact factor (2016):* 2.806. *Rank (Medicine (miscellaneous), 2016):* 245/1806 (Q1).

- Francisco Gómez-García[*]; Juan Ruano[*]; Macarena Aguilar-Luque; Patricia Alcalde-Mellado; Jesús Gay-Mimbrera; José Luis Hernández-Romero; Juan Luis Sanz-Cabanillas; Beatriz Maestre-López; Marcelino González-Padilla; Pedro Jesús Carmona-Fernández; Antonio Vélez García-Nieto; Beatriz Isla-Tejera. **Abstract analysis method facilitates filtering low-methodological quality and high-bias risk systematic reviews on psoriasis interventions.** *BMC Medical Research Methodology* 2017 [*in press*]
  Índices bibliométricos: *Journal Impact factor (2016):* 3.295. *Rank (Epidemiology, 2016):* 17/86 (Q1).

- J. Ruano[*]; M. Aguilar-Luque[*]; F. Gómez-García; P. Alcalde-Mellado; J. Gay-Mimbrera; P. J. Carmona-Fernandez; B. Maestre-Lopez; J.L. Sanz-Cabanillas; J.L. Hernandez-Romero; M. González-Padilla; A. Vélez García-Nieto; B. Isla-Tejera.**The differential impact of scientific quality, bibliometric factors, and social media activity on the influence of systematic reviews and meta-analyses about psoriasis**. *PLoS One* 2017 [*in press*]
  Índices bibliométricos: *Journal Impact factor (2016):* 2.806. *Rank (Medicine (miscellaneous), 2016):* 245/1806 (Q1).

[*] Igual contribución.

British Journal of Dermatology

Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.

Francisco Gomez-Garcia[1,2], David Epstein[3], Beatriz Isla-Tejera [2,4], Ana Lorente[1,2], Antonio Velez Garcia-Nieto[1,2], Juan Ruano[1,2,*].

[1]Department of Dermatology, Reina Sofía University Hospital, 14004 Córdoba, Spain.
[2]IMIBIC/Reina Sofía University Hospital/University of Cordoba, 14004 Córdoba, Spain.
[3]Department of Pharmacy, Reina Sofía University Hospital, 14004 Córdoba, Spain.
[4]Department of Applied Economics, School of Economics, University of Granada, Granada, Spain

[*]**Corresponding author**. Tel.: +34 9 57011213; fax: +34 9 57011214.
E-mail: juanruanoruiz@mac.com

Funding sources: This work was supported, in part, by project ICI1400136 to JR, integrated into the National Plan of R+D+I 2008-2011 and cofinanced by the ISCIII-Subdirección General de Evaluación and European Regional Development Fund (ERDF), and by grant PP13/009 of Plan Propio de movilidad para investigadores del Instituto Maimonides de Investigacion Biomédica de Córdoba (IMIBIC). No funding was received from any pharmaceutical company.

Conflict of Interest: J.R. has received honoraria for lecturing and grants for research from Pfizer, honoraria for lecturing from Janssen-Cilag and Novartis, and other financial benefits from AbbVie and Novartis; A.V.G.N. has received honoraria for lecturing from Pfizer, Novartis, AbbVie and Janssen-Cilag, and other financial benefits from AbbVie, Novartis and Janssen-Cilag; F.G.G. has received honoraria for research from Pfizer, and for lecturing from AbbVie and Janssen-Cilag. A.L. has received honoraria for lecturing from AbbVie and other financial benefit from Novartis. D.E. has received funding from Boehringer Ingelheim and Roche. B.I.T. has no disclosures.

Gómez-García F, Epstein D, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis. Br J Dermatol. 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

Abstract

A new generation of biologics targeting the interleukin-23–T helper 17 pathway has been developed. This study aimed to assess the short-term effectiveness and safety of these new agents using a network meta-analysis. Twenty-seven randomized clinical trials (10,629 patients) were identified by a comprehensive systematic literature review (PROSPERO 2015: CRD42015025472). Quality of evidence was assessed following Cochrane-compliant rules and the Grading of Recommendations, Assessment, Development and Evaluations approach. Efficacy and safety outcomes at weeks 10-16 were compared using a random-effects network meta-analysis within a frequentist framework to estimate pooled odds ratios (ORs) of direct and indirect comparisons among the therapeutic options. There were six direct drug-to-drug comparisons in the network, with a high degree of consistency between the direct and indirect evidence. From the available evidence, infliximab 5 mg.kg-1 every 8 weeks [OR 118.89, 95% confidence interval (CI) 60.91-232.04] and secukinumab 300 mg every 4 weeks (OR 87.07, 95% CI 55.01-137.82) are shown to be among the most effective short-term treatments, but are ranked as the biologics most likely to produce any adverse event or an infectious adverse event, respectively. Ustekinumab 90 mg every 12 weeks, the third most efficacious treatment (OR 73.67, 95% CI 46.97-115.56), was the only agent that did not show increased risk of adverse events compared with placebo. Treatment recommendations should also consider long-term outcomes and costs.

Keywords: psoriasis; methodological quality; AMSTAR; systematic review; meta- analysis.

Introduction

Psoriasis is a chronic inflammatory skin disease mediated by the cells and components of both the innate and adaptive immune systems, affecting 1–3% of the general population.[1] Tumour necrosis factor (TNF)-α antagonists have been at the centre of treatment for patients with moderate-to-severe plaque psoriasis who are unresponsive to, or intolerant of, nonbiological systemic agents.[2,3] However, our understanding of this disease has progressed greatly, and new drugs targeting the cytokines involved in the interleukin (IL)-23-T helper 17 pathway have emerged.[4] Several studies have shown that a new generation of monoclonal antibodies that block IL-12/ 23p40, IL-23p19, IL-17A or IL-17RA can reverse the clinical, histological and molecular features of psoriasis in approximately 70–80% of patients, compared with 45–50% in the case of TNF-α antagonists.[5-7]

As patients with psoriasis require lifelong treatment, reliable evidence of the comparative benefits and harms of interventions is needed to make clinical decisions regarding their use. Meta-analyses are conducted to assess the strength of recommendations and quality of evidence available for a disease and multiple treatment alternatives, improving the precision of estimates of effect and answering questions not posed by the individual studies. In network meta-analyses (NMAs), several treatments can be compared by connecting evidence from clinical trials that have investigated two or more treatments. The resulting trial network may allow estimation of the relative effects of all pairs of treatments, taking direct and indirect evidence into account.[8] For this reason, they are gaining popularity among clinicians, guideline developers and health technology agencies as evidence on new interventions. Systematic review authors and assessors are now strongly encouraged to make use of the Preferred Reporting Items for Systematic review and Meta-analysis Protocol (PRISMA-P) when drafting and appraising review protocols.

Five NMAs have compared the short-term efficacy of treatments for moderate-to-severe psoriasis. Woolacott *et al.* and Bansback *et al.* compared the efficacy of systemic treatments for moderate-to-severe plaque psoriasis including anti-TNF-α agents such as etanercept, adalimumab or infliximab with that of other biologics and nonbiologics.[9,10] Recently, three studies included ustekinumab, an anti-IL-12/23p40 agent, in the NMA.[11,13] Signorovitch *et al.* used NMA to compare the efficacy of biological treatments for moderate-to-severe psoriasis, adjusting the model for placebo responses.[14] A systematic review and meta-analysis performed by Nast *et al.* evaluated direct evidence of the long-term efficacy and safety of some of these drugs.[13] Only the last two were performed following the PRISMA-P statement, but none used the recently published PRISMA NMA checklist.[15]

This study aims to extend the existing NMAs to assess the direct and indirect evidence for the short-term efficacy and safety of new biologics targeting the IL-23-T helper 17 pathway in comparison with anti-TNF-α drugs for the treatment of moderate-to-severe plaque psoriasis.

Materials and methods

Systematic review

This systematic review and meta-analysis was conducted in accordance with the modified 32-item PRISMA extension statement for NMA of 2015.[15] The selection of databases, eligibility criteria, outcomes of the review and analytical methods were defined *a priori* in an internal protocol and registered on PROSPERO under the code CRD42015025472 (Table S1; see Supporting Information). Interventional studies were eligible for inclusion if they were randomized, placebo-controlled or head-to-head trials published in English, of infliximab [5 mg kg 1 at weeks 0, 2 and 6 then every 8 weeks (Q8W)], etanercept [50 mg twice weekly (BIW) for 12 weeks, then 25 mg BIW or 50 mg Q1W], adalimumab (80 mg at week 0, 40 mg at week 1, then 40 mg Q2W), ustekinumab (45 mg or 90 mg at weeks 0 and 4, then Q12W) or secukinumab (300 mg at weeks 0, 1, 2 and 3, then Q4W) as monotherapy for the treatment of plaque psoriasis in adult patients.

Databases searched and study identification

Details regarding the databases searched and study identification for this review are provided in Appendix S1 (see Supporting Information).

Data extraction

Treatment effects were evaluated based on the intention-to-treat efficacy rates (PASI 75 and PASI 90; number of patients with five-point Investigator's Global Assessment, Physician's Global Assessment or static Physician's Global Assessment of 0 or 1 at weeks 10–16; number of patients with Dermatology Life Quality Index of 0 or 1 at weeks 10–16) and safety parameters [number of patients with at least one adverse event (AE), number of patients with at least one serious AE (SAE), number of patients with at least one infectious AE, and number of patients withdrawing owing to AE] reported in the randomized trials identified during the systematic review. PASI 75 and PASI 90 represent $>75\%$ and $>90\%$

reduction, respectively, in the PASI score with respect to baseline. For more details, see the PROSPERO register file and PRISMA NMA 2015 checklist (Tables S1, S2; see Supporting Information).

Quality of evidence and risk-of-bias assessment

Details regarding the quality of evidence and risk-of-bias assessment for this review are provided in Appendix S1 (see Supporting Information).[16]

Meta-analysis of direct treatment effects

Data extracted from trials were combined by a random-effects model, with effect sizes expressed as the odds ratio (OR) of achieving each outcome at weeks 10–16 in the treatment arm vs. the control arm. Total effect size was calculated by the Mantel–Haenszel method. Heterogeneity was evaluated with I^2 calculations. Statistical analysis was performed with RevMan 5.3* , with two-tailed P-values < 0.05 considered significant. Forest plots and funnel plots were obtained for each outcome analysed at 10–16 weeks.

Network meta-analysis

NMA was used to make mixed comparisons among the therapeutic options and to rank the treatments, using the package *mvmeta* in Stata version 12.0 (Stata Corp, College Station, TX, U.S.A.) (Appendix S2; see Supporting Information). Inconsistency between direct and indirect evidence in the network was analysed using the ratio of odds ratios. A value >2 indicates that the direct and indirect treatment comparisons may be inconsistent.[17]

Results

Results of the search

During the abstract review phase, 2,025 records were identified (Fig. S1; see Supporting Information). Twenty-seven studies assessed for eligibility met the inclusion criteria and were included in the qualitative and quantitative analysis (10,629 randomized patients: 6,540 to biologics and 4,089 to conventional treatments) (Table S3; see Supporting Information).[18-42]

*http://tech.cochrane.org/revman

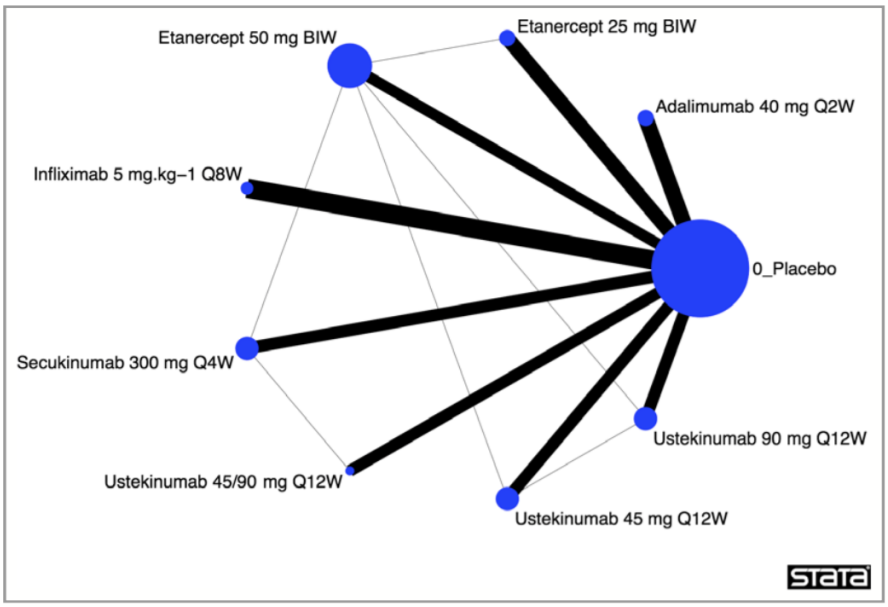


Fig. 1 **Network graph.** Treatments are represented by nodes, and head-to-head studies between treatments are represented by lines. The area of the node circle is proportional to the number of studies including that treatment, and the width of the line is proportional to the average effectiveness in the placebo arms of the studies. There were 15 direct comparisons in the network: nine of biologics vs. placebo and six of biologics against other biologics. The graph contains three loops and three closed loops. Both anti-interleukin-23-T helper 17 and antitumour necrosis factor-α agents are represented in them. Etanercept 50 mg twice a week (BIW) is the node with the highest connectivity (excluding placebo). Infliximab and adalimumab, two antitumour necrosis factor-α agents, are not present in any direct comparison with other biologic. Q2W, every 2 weeks.

There are only three included head-to-head trials reporting efficacy and safety data. The study sample size varied from 33 to 1,230. For the majority of trials, the double-blind period comprised 12 weeks (range 10-16). The mean prior disease duration was 18.3 years (range 10-23) and the disease severity was evaluated across the trials with a baseline PASI score of 20.2 (range 15-27) and a body surface area of 28.1% (range 20-45). Psoriatic arthritis was diagnosed in 27.3% of cases (range 14-78). Trials included 67% men (range 54-79). The mean age was 45 years (range 35-51) and the mean weight 90 kg (range 79-99). The majority of the patients were white. Detailed information for all of the included studies is presented in Table S3. A network graph summarizing the comparisons is provided in Figure 1.

Risk of bias

The risk of bias among the included studies was rated as 'low risk' or 'unclear risk' (Fig. S2; see Supporting Information). Of the 27 included clinical trials, 21 (78%) reported an adequate randomization method, and allocation concealment was properly ensured in 18 studies (67%). In all studies, the blinding of participants and personnel was sufficient. In 21 trials (78%), the risk of attrition bias was low, as incomplete outcome data were sufficiently addressed. The risk of reporting bias was low in most of the studies (78%).

Quality of evidence

When considering efficacy parameters, trials were of generally moderate quality according to GRADEpro assessment (Table S4; see Supporting Information). However, there was a wide range of values related to quality assessment of safety variables. Studies comparing agents head-to-head showed better quality of evidence than those reporting placebo as the comparator (Fig. S3; see Supporting Information). This was less evident when the comparator was another biologic belonging to a different pharmaceutical company.

Table 2 Mixed treatment comparisons estimated using random-effects network meta-analysis for 75% and 90% improvement in Psoriasis Area and Severity Index (PASI) 75 and PASI 90 at weeks 10–16.

Treatment	Comp.	PASI 75	PASI 90	Direct comparison
Infliximab 5 mg.kg-1 Q8W	Placebo	118.9 (60.9-232.0)	84.1 (31.0-228.5)	5
Secukinumab 300 mg Q4W	Placebo	87.1 (55.0-137.8)	96.0 (48.8-188.6)	4
Ustekinumab 90 mg Q12W	Placebo	73.7 (47.0-115.6)	61.3 (13.1-28.7)	2
Ustekinumab 45 mg Q12W	Placebo	56.2 (36.0-87.8)	56.0 (20.6-152.2)	2
Adalimumab 40 mg Q2W	Placebo	30.7 (21.5-43.9)	22.1 (8.2-60.0)	4
Etanercept 50 mg BIW	Placebo	17.9 (14.0-22.9)	16.5 (9.8-28.0)	9
Etanercept 25 mg BIW	Placebo	16.1 (9.2-23.0)	15.1 (5.1-44.8)	4

Efficacy and safety of direct comparisons of monotherapy vs. placebo

All biologics showed superior efficacy over placebo with respect to all efficacy outcomes (Table 1; Fig. 2; Fig. S4.2, 4.4, 4.6, 4.8-12; see Supporting Information). Secukinumab 300 mg Q4W and infliximab 5 mg.kg-1 Q8W were the most effective biologics, as demonstrated by their PASI 75 and PASI 90 responses.

All treatments showed a higher OR for 'at least one AE', although the ORs for adalimumab and for all doses of ustekinumab were not statistically significant (Fig. S4.9). Compared with placebo, no significant differences in the risk of 'at least one infectious

AE' were shown for etanercept 25 mg BIW, ustekinumab 45 mg Q12W or ustekinumab 90 mg Q12W (Fig. S4.10). After short-term treatment, no significant differences were observed in the risks of 'at least one SAE' for any agent (Fig. S4. 11). When considering 'withdrawal due to AE', no significant risk differences were found, except for ustekinumab 90 mg Q12W, which showed a lower risk (OR 0.14, 95%CI 0.03-0.63) (Fig. S4.12).

Table 3 Results of pooled odds ratios of the direct comparisons of each biologic included in the network vs. placebo

Treatment 1	Treatment 2	PASI 75	PASI 90
Adalimumab 40 mg Q2W	Etanercept 25 mg BIW	2.52 (1.46–4.35)	2.49 (0.95–6.52)
	Etanercept 50 mg BIW	1.44 (0.92–2.26)	1.23 (0.55-2.73
	Infliximab 5 mg.kg.-1 Q8W	0.26 (0.12–0.56)	0.26 (0.07-0.92)
	Secukinumab 300 mg Q4W	0.33 (0.20–0.55)	0.25 (0.10-0.65)
	Ustekinumab 45 mg/90 mg Q12W	0.74 (0.43–1.26)	0.53 (0.21-1.31)
	Ustekinumab 45 mg Q12W	0.90 (0.50–1.62)	0.70 (0.22-2.26)
	Ustekinumab 90 mg Q12W	0.57 (0.34–0.98)	0.44 (0.18-1.10)
Etanercept 25 mg BIW	Etanercept 50 mg BIW	0.57 (0.41–0.79)	0.49 (0.28-0.87)
	Infliximab 5 mg.kg-1 Q8W	0.10 (0.05-0.22)	0.10 (0.03-0.37)
	Secukinumab 300 mg Q4W	0.13 (0.08-0.021)	0.10 (0.05-0.23)
	Ustekinumab 45 mg Q12W	0.29 (0.19-0.46)	0.21 (0.10-0.46)
	Ustekinumab 45 mg/90 mg Q12W	0.36 (0.21-0.62)	0.28 (0.10-0.82)
	Ustekinumab 90 mg Q12W	0.23 (0.15-0.33)	0.18 (0.08-0.39)
Infliximab 5 mg.kg-1 Q8W	Secukinumab 300 mg Q4W	1.30 (0.61-2.75)	0.98 (0.28-3.39)
	Ustekinumab 45 mg Q12W	2.89 (1 34–6 22)	2.02 (0.60–6.85)
	Ustekinumab 45 mg/90 mg Q12W	3.54 (1.59–7.90)	2.71 (0.65–11.21)
	Ustekinumab 90 mg Q12W	2.25 (1 05–4 83)	1.71 (0.51–5.77)
Secukinumab 300 mg Q4W	Ustekinumab 45 mg Q12W	2.23 (1.44–3.47)	2.07 (0.95–4.54)
	Ustekinumab 45 mg/90 mg Q12W	2.73 (1.79–4.18)	2.77 (1.40–5.51)
	Ustekinumab 90 mg Q12W	1.74 (1.13–2.68)	1.75 (0.80–3.82)
Ustekinumab 45 mg Q12W	Ustekinumab 45 mg/90 mg	1.22 (0.72–2.10)	1.34 (0.47–3.79)
	Ustekinumab 90 mg Q12W	0.78 (0.62–0.99)	0.84 (0.57–1.25)
Ustekinumab 45 mg/90 mg Q12W	Ustekinumab 90 mg Q12W	0.64 (0.38–1.08)	0.63 (0.22–1.78)

Mixed treatment comparisons

When considering PASI 75 response, adalimumab 40 mg Q2W is 2.5-fold more effective than etanercept 25 mg BIW (OR 2.52, 95% CI 1.46-4.35), while infliximab 5 mg.kg-1 Q8W

(OR 0.26, 95% CI 0.12-0.56) and secukinumab 300 mg Q4W (OR 0.33, 95% CI 0.20-0.55) are about four- and threefold more effective than adalimumab 40 mg Q2W, respectively (Table 2).

Inconsistency analysis

The indirect comparison of etanercept 50 mg BIW vs. ustekinumab 45 mg Q12W or ustekinumab 90 mg Q12W showed a treatment effect about twice as large as the direct evidence (Table 3). The direct evidence (comparing one biologic with another head to head) does not show any significant differences between the treatments in any AE (Fig. S5; see Supporting Information). However, as already noted, use of secukinumab 300 mg Q4W carries a significantly greater risk of infectious AEs than etanercept 50 mg BIW (OR 1.64, 95% CI 1.04–2.59) and ustekinumab 45 mg or 90 mg Q12W (OR 2.56, 95% CI 1.23–5.26).

Table 4 Pooled odds ratios of the indirect, direct and mixed treatment comparisons of each head-to-head analysis included in the network. This table summarizes the direct treatment comparisons for efficacy (75% and 90% improvement in Psoriasis Area and Severity Index, PASI 75 and PASI 90) and safety (patients with at least one adverse event or patients with at least one infectious adverse event), and shows the indirect and mixed comparisons calculated by the network meta-analysis. Comparison results are represented as: OR (95%CI).

Treatment 1	Treatment 2	Indirect comparisons	Direct comparisons	Mixed comparisons	RoR	P-value	n
PASI 75							
Secukinumab 300 mg Q4W	Etanercept 50 mg BIW	4.74 (2.92–7.70)	4.28 (3.05–6.02)	4.37 (3.16–6.05)	1.10	0.85	1
Ustekinumab 90 mg Q12W	Etanercept 50 mg BIW	4.12 (2.46–6.87)	2.15 (1.55–2.94)	2.50 (1.82–3.44)	1.92	0.034*	1
Ustekinumab 45 mg Q12W	Etanercept 50 mg BIW	3.14 (1.88–5.23)	1.57 (1.10–2.26)	1.95 (1.40–2.73)	1.99	0.030*	1
Ustekinumab 90 mg Q12W	Ustekinumab 45 mg Q12W	1.31 (0.69–2.47)	1.28 (0.97–1.69)	1.28 (1.01–1.62)	1.02	0.81	1
Ustekinumab 45 mg/90 mg Q12W	Secukinumab 300 mg Q4W	0.40 (0.20–0.79)	0.35 (0.21–0.58)	0.36 (0.24–0.56)	1.15	0.92	1
Etanercept 50 mg BIW	Etanercept 25 mg BIW	1.11 (0.60–2.09)	1.89 (1.40–2.56)	1.75 (1.26–2.42)	1.63	0.11	2
PASI 90							
Secukinumab 300 mg Q4W	Etanercept 50 mg BIW	6.09 (2.37–15.68)	4.41 (3.19–6.39)	4.83 (2.59–8.98)	1.35	0.59	1

Ustekinumab 90 mg Q12W	Etanercept 50 mg BIW	3.46 (1.26–9.44)	2.69 (1.94–3.73)	2.75 (1.58–4.80)	1.28	0.78	1
Ustekinumab 45 mg Q12W	Etanercept 50 mg BIW	3.36 (1.23–9.17)	1.91 (1.31–2.78)	2.33 (1.32–4.10)	1.76	0.44	1
Ustekinumab 90 mg Q12W	Ustekinumab 45 mg Q12W	1.02 (0.30–3.50)	1.18 (0.84–1.67)	1.18 (0.78–1.77)	1.15	0.61	1
Ustekinumab 45 mg/90 mg Q12W	Secukinumab 300 mg Q4W	–	0.36 (0.25–0.50)	0.36 (0.17–0.73)	–	–	1
Etanercept 50 mg BIW	Etanercept 25 mg BIW	1.08 (0.31–3.75)	2.14 (1.41–3.24)	2.03 (1.15–3.58)	1.98	0.18	2
Any adverse event							
Secukinumab 300 mg Q4W	Etanercept 50 mg BIW	1.01 (0.75–1.37)	0.96 (0.70–1.31)	1.02 (0.83–1.26)	1.05	0.85	1
Ustekinumab 90 mg Q12W	Etanercept 50 mg BIW	0.75 (0.55–1.03)	0.96 (0.69–1.33)	0.80 (0.65–0.99)	1.28	0.23	1
Ustekinumab 45 mg Q12W	Etanercept 50 mg BIW	0.95 (0.69–1.30)	0.83 (0.58–1.20)	0.91 (0.73–1.13))	1.14	0.61	1
Ustekinumab 90 mg Q12W	Ustekinumab 45 mg Q12W	0.80 (0.54–1.17)	0.88 (0.70–1.11)	0.88 (0.73–1.07)	1.1	0.13	1
Ustekinumab 45 mg per 90 mg Q12W	Secukinumab 300 mg Q4W	0.92 (0.57–1.46)	0.78 (0.57–1.07)	0.84 (0.65–1.07)	1.09	0.50	1
Etanercept 50 mg BIW	Etanercept 25 mg BIW	0.58 (0.37–0.90)	1.05 (0.70–1.57)	0.73 (0.51–1.03)	1.81	0.10	2
Infectious adverse event							
Secukinumab 300 mg Q4W	Etanercept 50 mg BIW	1.64 (1.04–2.59)	1.15 (0.81–1.64)	1.38 (0.98–1.96)	1.42	0.30	1
Ustekinumab 90 mg Q12W	Etanercept 50 mg BIW	0.83 (0.51–1.35)	1.03 (0.74–1.43)	0.90 (0.65–1.26)	1.24	0.67	1
Ustekinumab 45 mg Q12W	Etanercept 50 mg BIW	0.99 (0.62–1.60)	1.08 (0.74–1.56)	1.03 (0.73–1.44))	1.09	0.90	1
Ustekinumab 90 mg Q12W	Ustekinumab 45 mg Q12W	0.84 (0.48–1.45)	0.88 (0.66–1.17)	0.88 (0.65–1.18)	1.05	0.66	1
Ustekinumab 45 mg/90 mg Q12W	Secukinumab 300 mg Q4W	0.39 (0.19–0.81)	0.82 (0.58–1.15)	0.65 (0.43–0.99)	2.10	0.029*	1
Etanercept 50 mg BIW	Etanercept 25 mg BIW	1.04 (0.59–1.82)	0.97 (0.54–1.74)	0.99 (0.62–1.59)	1.07	0.93	2

Ranking of treatments by efficacy

Infliximab 5 mg kg 1 Q8W and secukinumab 300 mg Q4W are ranked the most effective (PASI 75 and PASI 90, respectively), followed by ustekinumab 90 mg Q12W (Table 4). The probability that secukinumab 300 mg Q4W will be the most effective treatment option (PASI 90) for particular cases is almost 46 9%, compared with infliximab 5 mg kg 1 Q8W (44.6%) and ustekinumab 90 mg Q12W (4.3%) (Fig S6; see Supporting Information).

Ranking of treatments by safety

Etanercept 25 mg BIW ranked most likely to produce any AE (Table 4). This arises principally from a single trial,[26] which estimated an OR of 4.59 (95% CI 1.88-11.20) compared with placebo (Figs S4 9, S7; see Supporting Information). It seems clinically implausible that a lower dose would produce considerably more AEs than a higher dose, and this result may be due to a different patient population or methods of outcome measurement. Among the other treatments, infliximab 5 mg.kg-1 Q8W is ranked the second highest for risk of any AE. Secukinumab 300 mg Q4W is ranked as the treatment most likely to produce an infectious AE (OR 2.05, 95% CI 1.57-2.67) (Fig. S4 11).

Table 5 Ranking by the effectiveness and safety of the treatments included in the network

Treatment	SUCRA	PrBest	Mean rank
PASI 75 at weeks 10-16			
Infliximab 5 mg.kg-1 Q8W	95.9	70.8	1.3
Secukinumab 300 mg Q4W	90.9	28.9	1.7
Ustekinumab 90 mg Q12W	73.6	0.3	3.1
Ustekinumab 45 mg Q12W	58.3	0	4.3
Ustekinumab 45 mg/90 mg Q12W	48.9	0	5.1
Adalimumab 40 mg Q2W	43.2	0	5.5
Etanercept 50 mg BIW	26.6	0	6.9
Etanercept 25 mg BIW	12.6	0	8
Placebo	0	0	9
PASI 90 at weeks 10-16			
Infliximab 5 mg.kg-1 Q8W	86.6	46.2	2.1
Secukinumab 300 mg Q4W	91	46.7	1.7
Ustekinumab 90 mg Q12W	71.7	4.3	3.3
Ustekinumab 45 mg Q12W	63.4	1.9	3.9
Ustekinumab 45 mg/90 mg Q12W	52.9	0.5	4.8
Adalimumab 40 mg Q2W	39.7	0.4	5.8
Etanercept 50 mg BIW	30.6	0	6.6
Etanercept 25 mg BIW	14.1	0	7.9
Placebo	0	9	5.7

Patients with at least one AE			
Infliximab 5 mg.kg-1 Q8W	90	42.1	1.8
Secukinumab 300 mg Q4W	64.9	1.3	3.8
Ustekinumab 90 mg Q12W	21.2	0	7.3
Ustekinumab 45 mg Q12W	43.8	0.1	4.5
Ustekinumab 45 mg/90 mg Q12W	33.1	0	6.4
Adalimumab 40 mg Q2W	37.4	0.3	6
Etanercept 50 mg BIW	61.4	0.4	4.1
Etanercept 25 mg BIW	92.4	55.8	1.6
Placebo	5.7	0	8.5
Patients with at least one infectious AE			
Infliximab 5 mg.kg-1 Q8W	29.2	1.3	6.7
Secukinumab 300 mg Q4W	87.9	52.2	2
Ustekinumab 90 mg Q12W	39.1	1.1	5.9
Ustekinumab 45 mg Q12W	58.8	6.7	4.3
Ustekinumab 45 mg/90 mg Q12W	38	2.5	6
Adalimumab 40 mg Q2W	71.8	22.3	3.3
Etanercept 50 mg BIW	52.6	3.9	4.8
Etanercept 25 mg BIW	52.5	10	4.8
Placebo	20	0	7.4

Efficacy vs. safety at weeks 10–16 according to treatments and targeting pathways.

Infliximab 5 mg.kg-1 Q8W is shown to be among the most effective treatments in terms of PASI 75 response, but is ranked as the biologic most likely to produce an AE. However, infliximab 5 mg.kg-1 Q8W is ranked as one of the treatments least likely to produce any infectious AEs. On the other hand, secukinumab 300 mg Q4W is also among the most effective treatments in terms of PASI 90 response, but is ranked most likely to produce an infectious AE.

Table 6 Results of pooled odds ratios of the direct comparisons of each biologic included in the network vs. placebo.

Treatment	Comparator	PASI 75	PASI 90	Trials
Infliximab 5 mg.kg-1 Q8W	Placebo	118.9 (60.9–232.0)	84.1 (31.0–228.5)	5
Secukinumab 300 mg Q4W	Placebo	87.1 (55.0–137.8)	96.0 (48.8–188.6)	4
Ustekinumab 90 mg Q12W	Placebo	73.7 (47.0–115.6)	61.3 (13.1–287)	2
Ustekinumab 45 mg Q12W	Placebo	56.2 (36.0–87 8)	56.0 (20 6–152 2)	2

Values are the odds ratio (95% confidence interval). PASI 75, 75% improvement in Psoriasis Area and Severity Index; BIW, twice a week; Q2W, every 2 weeks. *Number of trials making direct comparison

Adalimumab 40 mg Q2W	Placebo	30.7 (21.5–43.9)	22.1 (8.2–60.0)	4
Etanercept 50 mg BIW	Placebo	17.9 (14.0–22.9)	16.5 (9.8–28.0)	9
Etanercept 25 mg BIW	Placebo	16.1 (9.2–23.0)	15.1 (5.1–44.8)	4

Discussion

This review is the first study to evaluate the short-term effects of secukinumab in patients with psoriasis, in addition to an agent that blocks IL-12/23p40 (ustekinumab) and the classical anti-TNF-α drugs (adalimumab, etanercept and infliximab). Our results suggest a division of treatments in terms of short- term effectiveness, with infliximab and secukinumab as the most effective agents and adalimumab and etanercept as the least effective options. The efficacy of ustekinumab is positioned at an intermediate point between those drugs, but with a higher overall quality of evidence and better safety profile than infliximab and secukinumab.

Given that head-to-head comparisons between biologics are scarce, one of the strengths of this study was the use of NMA. In the absence of trials involving a direct comparison of treatments of interest, an indirect comparison can provide useful evidence for the difference in treatment effects between competing interventions (which otherwise would be lacking) and for judiciously selecting the best choice(s) of treatment. For example, in our study we were able indirectly to compare infliximab 8 mg Q8W and adalimumab 40 mg Q2W against any other treatment, comparisons that have never been made in any randomized controlled trial.

Several factors typically need to be taken into account when recommending an intervention, not only its efficacy. Many systematic reviews therefore examine measures of both efficacy and safety, and the ranking of competing treatments for these two outcomes might differ considerably. In our study we decided to take into account both sources of information together to display the relative position of every treatment according to these variables. ORs for infectious AEs were higher for secukinumab 300 mg Q4W, adalimumab 40 mg Q2W, and etanercept 50 mg BIW than with placebo. In the majority of cases, infectious AEs were moderate or mild. Indeed, severe cases of infection were probably accounted for as SAEs, and are thus difficult to identify in most RCTs.

In the case of safety, the quality of the long-term studies evaluated by Nast et al. was considered moderate or low, with the quality of evidence strongly limited.[13] This situation is very similar to that which we found in relation to short-term safety. This highlights the low quality of evidence of AE-related published data provided by authors, which is in contrast to the efficacy of these drugs.

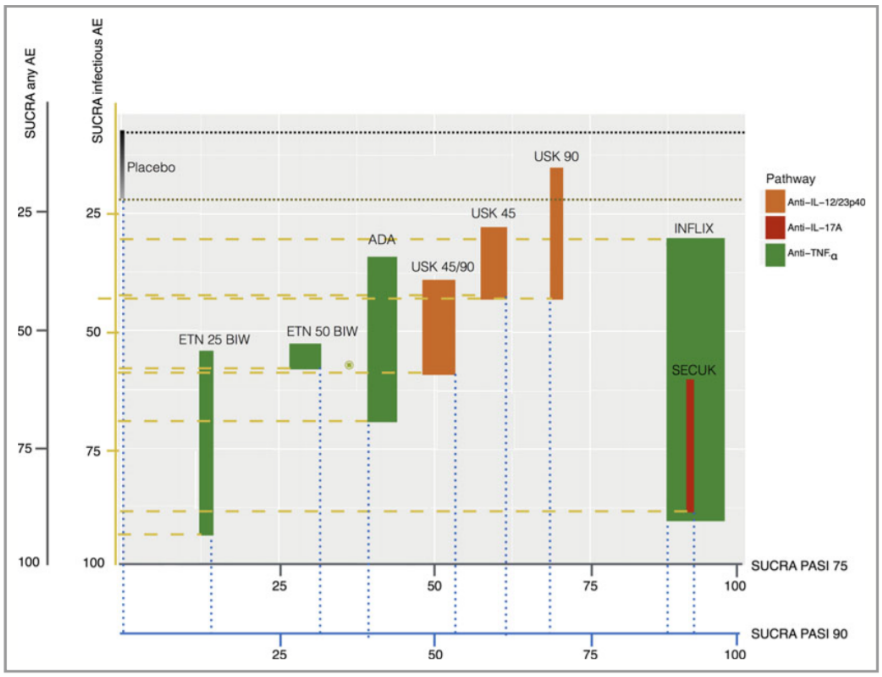


Fig. 2 Scatter plot of surface under the cumulative ranking (SUCRA) values for efficacy vs. safety outcomes at weeks 10–16 according to treatments and targeting pathways. This graph displays four separate sets of data using two y-axes for SUCRA of 75% improvement in Psoriasis Area and Severity Index (PASI 75) and PASI 90, and two x-axes for SUCRA of patients with at least one adverse event and patients with at least one infectious adverse event, based on the mixed treatment comparisons. Each treatment is represented by a geometric figure created by crossing the projection of these SUCRA values. Blue dotted lines represent SUCRA for PASI 90 values. Yellow dashed lines represent SUCRA for patients with at least one infectious adverse event. Points closer to the bottom left of the figure are relatively less effective and have fewer adverse events, while points closer to the top right are relatively more effective and relatively more at risk of adverse events. ADA, adalimumab; BIW, twice a week; ETN, etanercept; IL, interleukin; INFLIX, infliximab; SECUK, secukinumab; TNF, tumour necrosis factor; USK, ustekinumab.

Despite this limitation, the use of the PRISMA statement when drafting and appraising review protocols has improved the quality of the evidence from systematic reviews and meta-analyses. Recently, a modified 32-item PRISMA extension statement has been published for the reporting of systematic reviews incorporating NMAs.[15] This extension adds five new items that authors should consider when reporting an NMA, as well as 11 modifications to the existing PRISMA items. To the best of our knowledge no meta-analysis has previously evaluated the short-term efficacy of all of the approved drugs for psoriasis treatment, including secukinumab, based on these reporting guidelines.

Our study is limited to using PASI 75 and PASI 90 as the primary end points of the trials. Nevertheless, the lack of PASI 90 values in older clinical trials makes it difficult to compare them with new ones. Reich et al. handled the missing data by jointly modelling PASI 50, PASI 75 and PASI 90 achievement in a Bayesian hierarchical framework.[12] Signorovitch *et al.* tried to address this problem by modelling the relative risk of PASI 90 score adjusted by the response rate in the placebo group in each trial, but this adjustment did not substantially change the results of the NMA.[14] We believe the more traditional analysis shown here, although restricted to PASI 75 and PASI 90, is more transparent and more robust as it requires fewer modelling assumptions.

Another limitation of our study is that response to treatment over such short time periods may not be representative of the long-term effects of treatment, and the side-effect profile of individual treatments may be an important determinant of long-term success. In most of the long-term studies, the placebo groups are discontinued after 10-16-weeks of induction, making it difficult to obtain indirect evidence by means of NMA beyond this period. Recently, Nast *et al.* found infliximab, secukinumab and ustekinumab, followed by adalimumab and etanercept, to be the most efficacious long-term treatments in patients with moderate-to-severe psoriasis,[13] which is in line with the results reported herein. They employed an imputation approach to obtain long-term efficacy data. This approach led to 'serious' or 'very serious' risk of bias and an overall quality-of-evidence score of 'low' for most of the efficacy outcomes.

One of the key elements of comparative treatment effectiveness research is head-to-head trials. In our review, network-of-treatment comparisons reflect a predominant use of placebo as a comparator, resulting in a lack of direct comparison of biologics, which represented only 40% of the direct comparisons in the network and 11% of the 27 RCTs reviewed in this study. Estellat *et al.* found similar results (27.8% and 9.5%, respectively) when they analysed randomized controlled trials of biologics for rheumatoid arthritis.[43]

Another potential limitation is the inconsistency in some comparisons. In our study, we found significant inconsistency between indirect and direct PASI 75 estimates

for etanercept 50 mg BIW vs. ustekinumab 45 mg Q12W or ustekinumab 90 mg Q12W comparisons. Inconsistency in the treatment effect might arise from heterogeneity in the underlying severity of disease in the populations across different studies. However, we did not find any differences in variables such as weight or ethnicity among agents or studies.

In our study, visual inspection of funnel plots revealed possible publication bias in many of the active vs. placebo comparisons. However, an asymmetrical funnel plot should not be equated with publication bias. In our case, many trials with different drugs were plotted in the same funnel plot. Taking into account all plotted studies as if they were related to the same drug is not appropriate, and it may be the reason for this visual bias. Nevertheless, there seems to be publication bias against null results for PASI 75 and PASI 90 outcomes for secukinumab trials compared with other agents.

Finally, this meta-analysis includes only all currently licensed psoriasis biologics. Some published phase II and phase III clinical trials have studied other biologics that block IL-23p19 (guselkumab,[44] tildrakizumab),[45] IL-17A (ixekizumab)[29] or IL-17RA (brodalumab)[42,46] and investigated the dose–response relationship for the drug. Biologics that are pending licensing were not included in this study for two reasons. Firstly, only approved drugs provide estimates of comparative effectiveness that will be potentially useful to current decision makers. Secondly, addressing the PROSPERO rules requires submitting the study protocol to the database at least 6 months before the anticipated completion date, and this will undoubtedly help to reduce unplanned duplication and increase transparency, helping safeguard against selective reporting.

In conclusion, from the available evidence, infliximab, secukinumab and ustekinumab were found to be the most efficacious short-term treatments for moderate-to-severe plaque psoriasis. However, infliximab and secukinumab showed the highest risk for any AEs and associated mild-to-moderate infections, respectively, while ustekinumab was the agent with the best efficacy–safety profile. Our results could potentially aid the future assessment of the incremental cost-effectiveness of alternative treatments, and may provide a useful basis for the preparation of treatment guidelines for the use of a new generation of biological therapies in moderate-to-severe plaque psoriasis. Further research is warranted to enable direct and indirect comparison of the efficacy and safety of psoriasis biological agents after long-term follow-up.

Acknowledgments

The authors would like to thank Editage (www.editage.com) for English-language editing.

References

[1] Lowes MA, Suarez-Farin as M, Krueger JG. Immunology of psoriasis. Annu Rev Immunol 2014; 32:227–55.
[2] Rustin MH. Long-term safety of biologics in the treatment of moderate-to-severe plaque psoriasis: review of current data. Br J Dermatol 2012; 167(Suppl. 3):3–11.
[3]Pathirana D, Ormerod AD, Saiag P et al. European S3-guidelines on the systemic treatment of psoriasis vulgaris. J Eur Acad Dermatol Venereol 2009; 23(Suppl. 2):1–70.
[4]Di Cesare A, Di Meglio P, Nestle FO. The IL-23/Th17 axis in the immunopathogenesis of psoriasis. J Invest Dermatol 2009; 129:1339–50.
[5]Croxtall JD. Ustekinumab: a review of its use in the manage- ment of moderate to severe plaque psoriasis. Drugs 2011; 71:1733–53.
[6] Levin AA, Gottlieb AB. Specific targeting of interleukin-23p19 as effective treatment for psoriasis. J Am Acad Dermatol 2014; 70: 555–61.
[7] Chiricozzi A, Krueger JG. IL-17 targeted therapies for psoriasis. Expert Opin Investig Drugs 2013; 22:993–1005.
[8]Bafeta A, Trinquart L, Seror R, Ravaud P. Reporting of results from network meta-analyses: methodological systematic review. BMJ 2014; 348:g1741.
[9]Woolacott N, Hawkins N, Mason A et al. Etanercept and efal- izumab for the treatment of psoriasis: a systematic review. Health Technol Assess 2006; 10:1–233.
[10] Bansback N, Sizto S, Sun H et al. Efficacy of systemic treatments for moderate to severe plaque psoriasis: systematic review and meta-analysis. Dermatology 2009; 219:209–18.
[11] Lin VW, Ringold S, Devine EB. Comparison of ustekinumab with other biological agents for the treatment of moderate to severe plaque psoriasis: a Bayesian network meta-analysis. Arch Dermatol 2012; 148:1403–10.
[12]Reich K, Burden AD, Eaton JN, Hawkins NS. Efficacy of biologics in the treatment of moderate to severe psoriasis: a network meta- analysis of randomized controlled trials. Br J Dermatol 2012; 166:179–88.
[13] Nast A, Jacobs A, Rosumeck S, Werner RN. Efficacy and safety of systemic long-term treatments for moderate-to-severe psoriasis – a systematic review and meta-analysis. J Invest Dermatol 2015; 135:2641–8.
[14] Signorovitch JE, Betts KA, Yan YS et al. Comparative efficacy of biological treatments for moderate-to-severe psoriasis: a network meta-analysis adjusting for cross-trial differences in reference arm response. Br J Dermatol 2015; 172:504–12.

[15] Hutton B, Salanti G, Caldwell DM et al. The PRISMA extension statement for reporting of systematic reviews incorporating net- work meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015; 162:777–84.
[16] The Cochrane Collaboration. Assessing risk of bias in included studies. Available at: http://methods.cochrane.org/bias/assessing- risk-bias-included-studies (last accessed 25 August 2016).
[17] Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numeri- cal summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011; 64:163–71.
[18] Chaudhari U, Romano P, Mulcahy LD et al. Efficacy and safety of infliximab monotherapy for plaque-type psoriasis: a randomised trial. Lancet 2001; 357:1842–7.
[19] Gottlieb AB, Evans R, Li S et al. Infliximab induction therapy for patients with severe plaque-type psoriasis: a randomized, double- blind, placebo-controlled trial. J Am Acad Dermatol 2004; 51: 534–42.
[20] Reich K, Nestle FO, Papp K et al. Infliximab induction and maintenance therapy for moderate-to-severe psoriasis: a phase III, multicentre, double-blind trial. Lancet 2005; 366:1367–74.
[21] Menter A, Feldman SR, Weinstein GD et al. A randomized compar- ison of continuous vs. intermittent infliximab maintenance regi- mens over 1 year in the treatment of moderate-to-severe plaque psoriasis. J Am Acad Dermatol 2007; 56(31):e1–15.
[2] Gottlieb AB, Matheson RT, Lowe N et al. A randomized trial of etanercept as monotherapy for psoriasis. Arch Dermatol 2003; 139:1627–32.
[23] Leonardi CL, Powers JL, Matheson RT et al. Etanercept as monotherapy in patients with psoriasis. N Engl J Med 2003; 349:2014–22.
[24] Papp KA, Tyring S, Lahfa M et al. A global phase III randomized controlled trial of etanercept in psoriasis: safety, efficacy, and effect of dose reduction. Br J Dermatol 2005; 152:1304–12.
[25] Tyring S, Gottlieb A, Papp K et al. Etanercept and clinical out- comes, fatigue, and depression in psoriasis: double-blind placebo- controlled randomised phase III trial. Lancet 2006; 367:29–35.
[26] van de Kerkhof PC, Segaert S, Lahfa M et al. Once weekly adminis- tration of etanercept 50 mg is efficacious and well tolerated in patients with moderate-to-severe plaque psoriasis: a randomized controlled trial with open-label extension. Br J Dermatol 2008; 159:1177–85.
[27] Strober BE, Crowley JJ, Yamauchi PS et al. Efficacy and safety results from a phase III, randomized controlled trial comparing the safety and efficacy of briakinumab with

etanercept and placebo in patients with moderate to severe chronic plaque psoriasis. Br J Dermatol 2011; 165:661–8.
[28] Gottlieb AB, Leonardi C, Kerdel F et al. Efficacy and safety of briakinumab vs. etanercept and placebo in patients with moderate to severe chronic plaque psoriasis. Br J Dermatol 2011; 165:652–60.
[29] Griffiths CE, Reich K, Lebwohl M et al. Comparison of ixekizumab with etanercept or placebo in moderate-to-severe psoriasis (UNCOVER-2 and UNCOVER-3): results from two phase 3 randomised trials. Lancet 2015; 386:541–51.
[30] Bachelez H, van de Kerkhof PC, Strohal R et al. Tofacitinib versus etanercept or placebo in moderate-to-severe chronic plaque psoriasis: a phase 3 randomised non-inferiority trial. Lancet 2015; 386:552–61.
[31] Gordon KB, Langley RG, Gottlieb AB et al. A phase III, randomized, controlled trial of the fully human IL-12/23 mAb briakinumab in moderate-to-severe psoriasis. J Invest Dermatol 2012; 132:304–14.
[32] Menter A, Tyring SK, Gordon K et al. Adalimumab therapy for moderate to severe psoriasis: a randomized, controlled phase III trial. J Am Acad Dermatol 2008; 58:106–15.
[33] Saurat JH, Stingl G, Dubertret L et al. Efficacy and safety results from the randomized controlled comparative study of adalimumab vs. methotrexate vs. placebo in patients with psoriasis (CHAM- PION). Br J Dermatol 2008; 158:558–66.
[34] Gordon KB, Duffin KC, Bissonnette R et al. A phase 2 trial of guselkumab versus adalimumab for plaque psoriasis. N Engl J Med 2015; 373:136–44.
[35] Leonardi CL, Kimball AB, Papp KA et al. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 76-week results from a randomised, dou- ble-blind, placebo-controlled trial (PHOENIX 1). Lancet 2008; 371:1665–74.
[36]Papp KA, Langley RG, Lebwohl M et al. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 52-week results from a randomised, dou- ble-blind, placebo-controlled trial (PHOENIX 2). Lancet 2008; 371:1675–84.
[37] Griffiths CE, Strober BE, van de Kerkhof P et al. Comparison of ustekinumab and etanercept for moderate-to-severe psoriasis. N Engl J Med 2010; 362:118–28.
[38] Langley RG, Elewski BE, Lebwohl M et al. Secukinumab in plaque psoriasis – results of two phase 3 trials. N Engl J Med 2014; 371:326–38.
[39] Blauvelt A, Prinz JC, Gottlieb AB et al. Secukinumab administration by prefilled syringe: efficacy, safety and usability results from a randomized controlled trial in psoriasis (FEATURE). Br J Dermatol 2015; 172:484–93.

[40] Paul C, Lacour JP, Tedremets L et al. Efficacy, safety and usability of secukinumab administration by autoinjector/pen in psoriasis: a randomized, controlled trial (JUNCTURE). J Eur Acad Dermatology Venereol 2015; 29:1082–90.
[41] Tha"ci D, Blauvelt A, Reich K et al. Secukinumab is superior to ustekinumab in clearing skin of subjects with moderate to severe plaque psoriasis: CLEAR, a randomized controlled trial. J Am Acad Dermatol 2015; 73:400–9.
[42] Lebwohl M, Strober B, Menter A et al. Phase 3 studies comparing brodalumab with ustekinumab in psoriasis. N Engl J Med 2015; 373:1318–28.
[43] Estellat C, Ravaud P. Lack of head-to-head trials and fair control arms: randomized controlled trials of biologic treatment for rheumatoid arthritis. Arch Intern Med 2012; 172:237–44.
[44] Sofen H, Smith S, Matheson RT et al. Guselkumab (an IL-23-speci- fic mAb) demonstrates clinical and molecular response in patients with moderate-to-severe psoriasis. J Allergy Clin Immunol 2014; 133:1032–40.
[45] Papp K, Tha"ci D, Reich K et al. Tildrakizumab (MK-3222), an anti- IL-23p19 monoclonal antibody, improves psoriasis in a phase 2b randomized placebo-controlled trial. Br J Dermatol 2015; 173:930–9.
[46] Papp KA, Leonardi C, Menter A et al. Brodalumab, an anti-interleukin-17-receptor antibody for psoriasis. N Engl J Med 2012; 366:1181–9.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:
Appendix S1. Supplementary materials and methods. Appendix S2. STATA script for NMA analysis.
Table S1. PROSPERO register file.
Table S2. PRISMA extension for NMA 2015 checklist.
Table S3. Baseline characteristics and results of each study. **Table S4.** Summary of finding tables (SoF).
Fig S1. Sample search strategy.
Fig S2. Risk of bias of the included trials.
Fig S3. Scoring heat map of the quality of evidence for each outcome across pooled studies.
Fig S4. Forest and funnel plots: Verum vs placebo.
Fig S5. Forest and funnel plots: Verum vs verum.

Fig S6. Surface Under the Cumulative RAnking curves (SUCRA) for all biologic treatments in the psoriasis network for efficacy outcomes.
Fig S7. Surface Under the Cumulative RAnking curves (SUCRA) for all biologic treatments in the psoriasis network for efficacy outcomes.

British Journal of Dermatology

Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest, and bibliometric indices as predictors of methodological quality

Francisco Gomez-Garcia[1,2,*], Juan Ruano[1,2,*], Macarena Aguilar-Luque[2], Jesus Gay-Mimbrera[2], Beatriz Maestre-López[2,3], J.L. Sanz-Cabanillas[1,3], Pedro J. Carmona-Fernandez[2], Marcelino Gonzalez-Padilla[1,2], Antonio Velez Garcia-Nieto[1,2], Beatriz Isla-Tejera[2,4].

[1] Department of Dermatology, Reina Sofía University Hospital, 14004 Córdoba, Spain.

[2]IMIBIC/Reina Sofía University Hospital/University of Cordoba, 14004 Córdoba, Spain.

[3] School of Medicine, University of Cordoba, 14004 Córdoba, Spain.

[4]Department of Pharmacy, Reina Sofía University Hospital, 14004 Córdoba, Spain.

Corresponding author. Tel.: +34 9 57011213; fax: +34 9 57011214.

E-mail: juanruanoruiz@mac.com

***Equal contributors.**

Funding sources: This work was supported, in part, by project ICI1400136 to JR, integrated into the National Plan of R+D+I 2008-2011 and co-financed by the ISCIII-Subdirección General de Evaluación and European Regional Development Fund (ERDF), and by grant PP13/009 of Plan Propio de movilidad para investigadores del Instituto Maimonides de Investigacion Biomédica de Córdoba (IMIBIC). No funding was received from any pharmaceutical company.

Conflict of Interest: FG-G has received honoraria for research from Pfizer, and for lecturing from AbbVie, Janssen-Cilag and Novartis; JR has received honoraria for lecturing and grants for research from Pfizer, honoraria for lecturing from Janssen-Cilag and Novartis, and other financial benefits from AbbVie and Novartis; AVG-N has received honoraria for lecturing from Pfizer, Novartis, AbbVie, and Janssen-Cilag, and other financial benefits from AbbVie, Novartis, and Janssen-Cilag. MA-L, JR, FG-G, and BI-T are members of the Cochrane Bias Methods Group and Skin Group. JLS-C, PA-M, BM-L, PJC-F, and MG-P have no disclosures.

Gómez-García F, Ruano J, Aguilar-Luque M, Gay-Mimbrera J, Maestre-Lopez B, Sanz-Cabanillas JL, Carmona-Fernández PJ, González-Padilla M, Vélez García-Nieto A, Isla-Tejera B. Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest and bibliometric indices as predictors of methodological quality. Br J Dermatol. 2017 Jun;176(6):1633-1644. doi: 10.1111/bjd.15380. Epub 2017 May 19.

Abstract

The quality of systematic reviews and meta-analyses about psoriasis remains unknown. This study aimed to assess the methodological quality of reviews published up to 2016 identified by a comprehensive systematic searching in MEDLINE, EMBASE, and the Cochrane Database (PROSPERO: CDR42016041611). The reviews' methodological quality was assessed by two raters who extracted information from full-articles. After total and per-item Assessment of Multiple Systematic Reviews (AMSTAR) scores were calculated, reviews were classified as low (0-4), medium (5-8), or high (9-11) quality. Article metadata and journal-related bibliometric indices were obtained. Principal component and multivariate ordinal logistic regression analyses were used to find predictors of methodological quality. Two hundred twenty studies were classified as high (17.2%), moderate (55%), or low (27.7%) quality. Lower compliance rates were found for question (Q) 5 (list of studies provided, 11.3%), Q10 (publication bias assessed, 27.8%), Q4 (status of publication included, 39.5%), and Q1 (a priori design provided, 41%) AMSTAR items. Factors such as meta-analysis included (odds ratio [OR], 6.21; 95% confidence interval [CI]: 2.78-14.85), funding by academic institutions (OR, 2.89; 95% CI, 1.11-7.89), article influence score (OR, 2.13; 95% CI, 1.05-6.67), 5- year impact factor (OR, 95% CI, 1.02-1.14), article page count (OR, 1.08; 95% CI, 1.02-1.15), and number of authors with a conflict of interest (OR, 0.9; 95% CI, 0.824- 0.985) significantly predicted a higher quality. The methodological quality of reviews published about psoriasis remains suboptimal. The type of funding sources and author disclosures may compromise study quality, increasing the risk of bias.

Keywords: psoriasis; methodological quality; AMSTAR; systematic review; meta-analysis.

Introduction

Moderate-to-severe forms of psoriasis are associated with significant comorbidity, impaired quality of life, and high direct and indirect costs. Therefore, the therapeutic decision-making process may include both clinical and economic factors.[1] Dermatologists, like many other physicians in their respective fields, often refer to systematic reviews (SRs) to guide their clinical decision making.

SRs have become the standard approach for the synthesis of primary studies in medical research. An SR uses systematic methods in each phase (i.e. identification, selection, risk assessment, analysis, and interpretation of results) to respond to a clearly formulated research question. SRs also use an objective search of the literature, apply predetermined inclusion and exclusion criteria and critically appraise what is found to be relevant. Meta-analyses (MAs) enable the quantitative synthesis of randomized controlled trials (RCTs) where appropriate.

However, reviews have limitations, and despite the care with which they are conducted SRs may yield different answers to the same question.[2] The value of an SR depends on its quality, which can be defined as the likelihood that the design will generate unbiased outcomes. Since 1988, more than 40 instruments have been developed to measure the quality of SRs.[2,3] The Assessment of Multiple Systematic Reviews (AMSTAR) checklist has been the most frequently used tool among various medical disciplines since its publication in 2007..[4] It is valid, reliable and feasible, and the total score is meaningful.[5,6] Although the original AMSTAR checklist was developed to assess reviews that included only RCTs, it can be applied to various SRs.[8] The AMSTAR checklist is now being used by numerous groups, including the Canadian Agency for Drugs and Technologies in Health, and the Cochrane Effective Practice and Organization of Care Group.8 However, it has not yet been used to evaluate the methodological quality of studies about psoriasis.

We hypothesized that scientific articles on psoriasis funded by the pharmaceutical industry or affiliated organizations would be more likely to have methodological bias than articles without industry-associated sponsorship. This study aimed to assess the methodological quality of SRs and MAs published on psoriasis, and to build a statistical model to predict the quality of any study based on the AMSTAR tool.

Materials and methods

Inclusion criteria

We included SRs or MAs published on skin psoriasis in scientific journals. Historical articles, abstracts of congresses, case reports, surveys, narrative reviews, narrative reports (i.e. those with a focus on understanding a concept), clinical practice guidelines, consensus documents, MAs performed without a systematic literature search, and reviews titled as literature reviews or integrative reviews were not included. Our retrieval was restricted to English-language reviews because of time limitations for project completion. There was no limitation on the year of publication or study population.

Search and selection methods

We established an a priori protocol and published it in the PROSPERO International Prospective Register of Systematic Reviews (CRD42016041611). Details regarding the search methods for identification and selection are provided in Appendix S1 (see Supporting Information). Lists of included and excluded studies are shown in Tables S1 and S2 (see Supporting Information).

Assessment of methodological quality

Two investigators (F.G.-G. and J.G.-M.) independently assessed the methodological quality of each review using the same data abstraction forms and 11-point AMSTAR criteria. We did not use the AMSTAR score as an inclusion criterion, but we identified and discussed differences in quality between reviews and used the review quality assessment to interpret the results when synthesized in this overview. Detailed information for the AMSTAR checklist and the system of rating the articles are presented in Table 1 and Table S3 (see Supporting Information).

Table 7 Percentage of compliance with individual items of the Assessment of Multiple Systematic Reviews (AMSTAR) checklist vs. AMSTAR score- based levels

	All articles, n = 220	Low quality, n = 61	Medium quality, n = 121	High quality, n = 38

Q1: Was an 'a priori' design provided?	90 (40.9)	12(20)	44(36.4)	34 (89)
Q2: Was there duplicate study selection and data extraction?	118 (53.6)	10 (16)	73 (60.3)	35 (92)
Q3: Was a comprehensive literature search performed?	192 (87.3)	37 (61)	117 (96.7)	38 (100)
Q4: Was the status of publications (i.e. grey literature) used as an inclusion criterion?	87 (39.5)	11 (18)	47 (38.8)	29 (76)
Q5: Was a list of studies (included and excluded) provided?	25 (11.4)	2 (3)	7 (0.6)	16 (42)
Q6: Were the characteristics of the included studies provided?	197 (89.5)	44 (72)	116 (95.9)	37 (97)
Q7: Was the scientific quality of the included studies assessed and documented?	137 (62.3)	10 (16)	89 (73.6)	38 (100)
Q8: Was the scientific quality of the included studies used appropriately in formulating conclusions?	131 (59.5)	8 (13)	86 (71.1)	37 (97)
Q9: Were the methods used to combine the findings of studies appropriate?	121 (55.0)	5 (8)	79 (65.3)	37 (97)
Q10: Was the likelihood of publication bias assessed?	61 (27.7)	2 (3)	37 (30.6)	22 (58)
Q11: Was the conflict of interest included?	152 (69.1)	35 (57)	81 (66.9)	36 (95)

Data extraction and analysis

For studies that fulfilled the inclusion criteria, we independently obtained metadata for every article, author and journal (Table S4; see Supporting Information). Ordinal logistic regression was used to assess the association between the dependent variable, AMSTAR-based quality level (high, moderate or low), and other article-related or journal-related independent variables. Studies were classified as Cochrane vs. non-Cochrane reviews based on the authors' affiliation to the Cochrane Collaborative Group.

Univariate logistic regression analysis was used to identify factors with potential influence on the study quality; factors with significance < 0.25 were entered into subsequent backward stepwise multivariate ordinal regression analysis to obtain the final prediction model. Model internal validity was assessed by the j cross-validation method using the 'caret' package in R (R Development Core Team; http://www.r-project.org). Principal component analysis (PCA) and hierarchical nonsupervised clustering were performed to discover potentially significant subgroups beyond the classical AMSTAR three-level classification. Radial plots were used to represent the median of the accomplishment frequency of each AMSTAR question among dermatology journals. AMSTAR total scores are summarized descriptively as the median and interquartile range. AMSTAR scores per

item are also summarized as the percentage of achievement when an analysis by journal was performed (i.e. radial plots).

We reported dichotomous data as odds ratios (ORs) with 95% confidence intervals (CIs). P-values < 0.05 were considered statistically significant. Graphs were produced and statistics were analysed using several packages of the R language. Our analysis can be fully reproduced using several source files containing raw data and R scripts stored at our GitHub hosting repository (https://github.com/info4cure/SRandMAQual ity Assessment).

Protocol vs. overview

Our planned search strategy recorded in PROSPERO was compared with the final reported review methods (Table S5; see Supporting Information). Our retrieval was restricted to English-language reviews because of time limitations for project completion. We did not add, omit or change outcomes after our protocol was published. No ethical approval was required for this study.

Results

Our database search yielded 1195 titles with potential relevance (699 in Embase and MEDLINE, 474 in Embase, 22 in MEDLINE and four also in the Cochrane Database). After excluding duplicate articles and screening the abstracts, 304 studies were eligible for full-text review. Thus, 220 reviews from 92 peer-reviewed journals were assessed (Fig. 1; Tables S1, S2; see Supporting Information).

General characteristics

In total, 741 authors from 520 different institutions and 32 countries published the included reviews. The median numbers of authors and institutions per review were 5 (range 2–20) and 3 (range 1–16), respectively. The author's H-index varied widely among studies, with the median being 14 (range 1–108). Half of the assessed studies were published in dermatology journals (54.5%, 120 of 220), with the largest portion published (83.6 %, 184 of 220) during the last 5 years. Most journals (63.2%, 139 of 220) were ranked as Q1 as defined by SCImago Journal & Country Rank (SJR), with a median impact factor of 3.029 (range 0.764–30,030). Only six (2.7%) reviews were authored by Cochrane researchers.

The frequency distribution of the research areas related to psoriasis is as follows: treatment (58. 6%, 129 of 220), comorbidities (21.4%, 47 of 220), pathogeny (14.5%, 32 of 220) and economical analysis (1.4%, three of 220). Fewer than half of the reviews enrolled

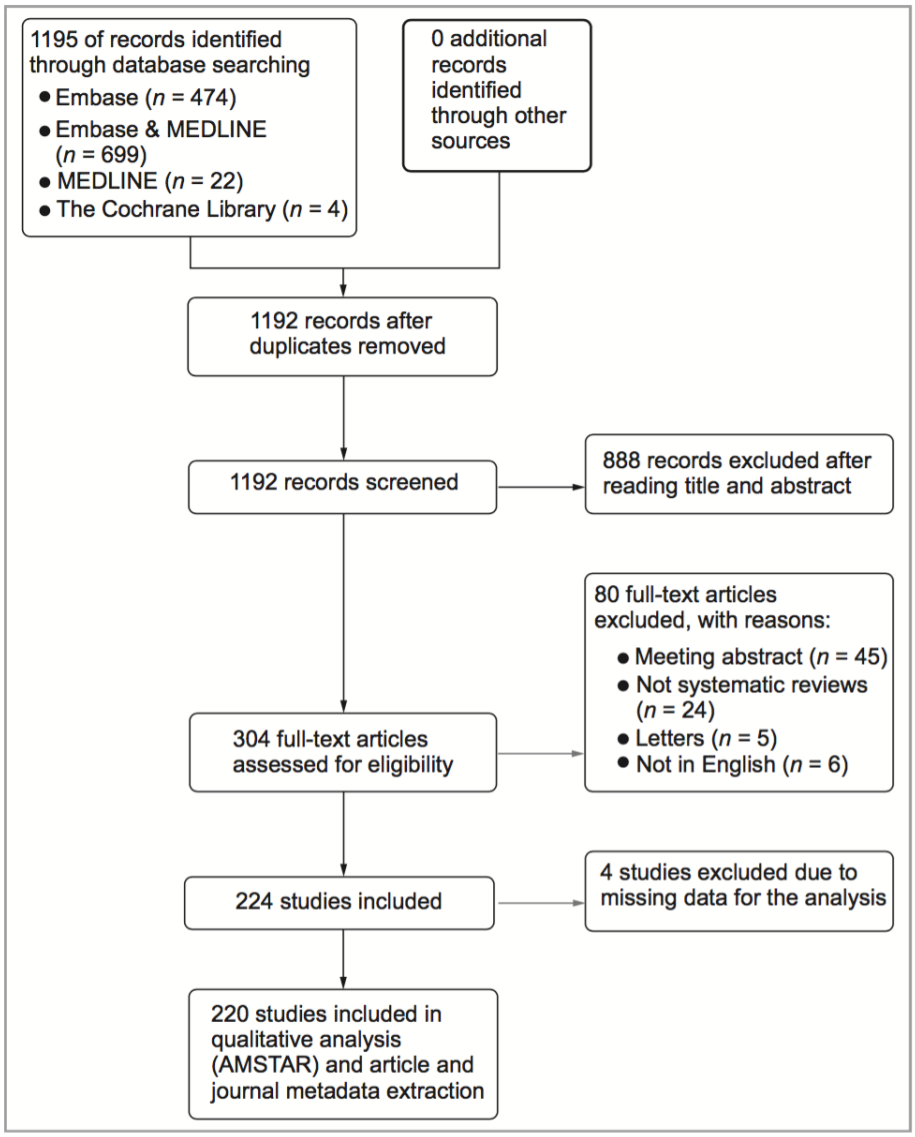


Fig. 1 PRISMA flowchart overview of systematic reviews and meta-analyses on psoriasis included in and excluded from the study. This figure shows the process description for selection and elimination of the studies at each stage of the systematic review.

primary studies with RCTs (43.2 %, 95 of 220), observational studies (36.8%, 81 of 220), RCTs and observational studies (5. 5%, 12 of 220), SRs (4.1%, nine of 220) and economic analysis (1.8%, four of 220).

Overall, 27.8% of reviews were funded by pharmaceutical companies, and 27.2% were funded by academic or health institutions. AbbVie Pharmaceuticals was the company that funded most of the reviews on psoriasis (79% of all funded studies), achieving a median AMSTAR score of 5.5 (range 1–10). Additionally, in 61.4% (135 of 220) of the included studies, authors declared a conflict of interest (in 50% of these cases, there were at least three authors with a conflict of interest per article).

The most prolific institutions, with 29 contributions, were the Department of Dermatology at Centre Hospitalier Universitaire de Brest (France), the Department of Dermatology at Hospital Edouard Herriot (France) and the Department of Dermatology at Radboud University Medical Centre (the Netherlands) (Table S6). The institutions with the highest median AMSTAR scores, although with a lower number of contributions and with more varied topics of research, were mainly from the U.K. and U.S.A., such as the Centre for Reviews and Dissemination and the Centre for Health Economics, both at the University of York (Table S7; see Supporting Information).

Assessment of methodological quality

Assessment of the methodological quality using AMSTAR questions began after agreement among reviewers became substantial (kappa = 0.75, 95% CI 0.69–0.82). The median AMSTAR score was 6 (interquartile range 4–8). Reviews were classified as high (17.2%), moderate (55.0%) or low (27.8%) quality. The AMSTAR items with the lowest compliance rates were: Q5 (‘list of studies provided’, 11.4%); Q10 (‘publication bias assessed’, 27.7%); Q4 (‘status of the publication included’, 39.5%); and Q1 (‘a priori design provided’, 40.9%). Q6, which assessed whether characteristics of the included studies were provided, had the best compliance rate in the AMSTAR checklist (89.5%). Total AMSTAR scores achieved by Cochrane Reviews were 10 or more. These studies represent only 15% of the high-quality methodological reviews subgroup.

Results

Our database search yielded 1195 titles with potential relevance (699 in Embase and MEDLINE, 474 in Embase, 22 in MEDLINE and four also in the Cochrane Database). After excluding duplicate articles and screening the abstracts, 304 studies were eligible

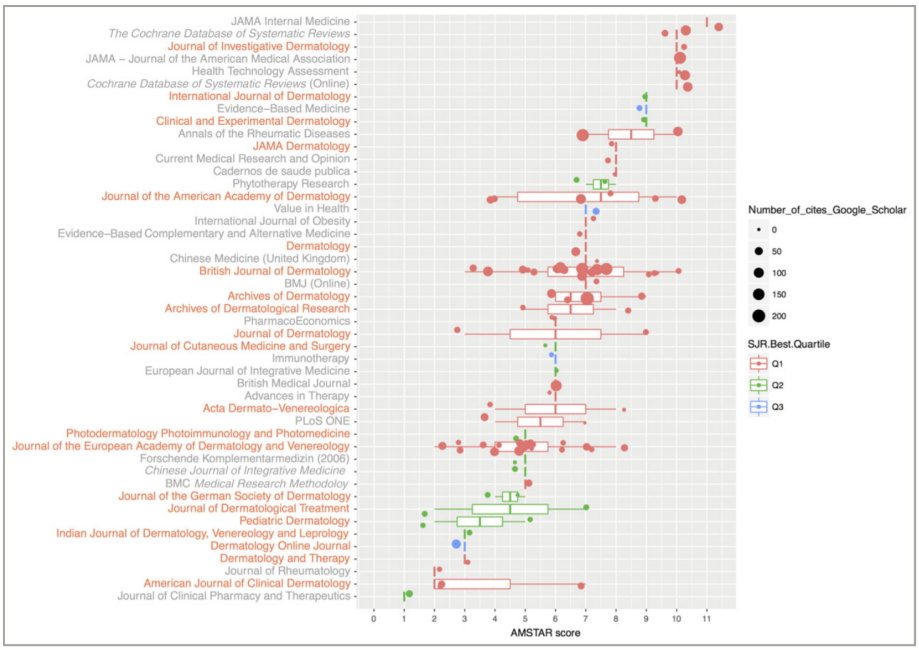


Fig. 2 Journal ranking based on the median Assessment of Multiple Systematic Reviews (AMSTAR) scores of published reviews. The bar plot displays AMSTAR scores of all reviews published per journal. Journals are ranked based on the median AMSTAR score per journal. Dermatology journals are highlighted in coral. Individual AMSTAR scores per article published in each journal are shown as points. The point size is proportional to the number of citations found in Google Scholar. Bar and point colours represent the journal's SCImago Journal Rank (SJR) best quartile.

for full-text review. Thus, 220 reviews from 92 peer-reviewed journals were assessed (Fig. 1; Tables S1, S2; see Supporting Information). Reviews were published in general medicine and dermatology journals with various bibliometric indices (Figs 2, 3). The most comprehensive and highest rated studies were a review related to the immunogenicity of antitumour necrosis factor-a agents, published in JAMA Internal Medicine by Maneiro et al.,[9] and an SR and MA of the efficacy and safety of topical treatments for scalp psoriasis published by Schlager et al. in the Cochrane Database of Systematic Reviews. [10]

Principal component analysis

We used PCA to convert the vectors of the 11 AMSTAR item subscores per article into a set of values of linearly uncorrelated variables called principal components (PCs). Figure 4(a) shows a PCA scatter plot of all included reviews on a coordinate system that optimally describes the variance between PC1 and PC2. With the objective of discovering new clusters of reviews based on methodological quality parameters, articles were ranked based on heat map clustering using the four most informative components of PCA (Fig. 4b–d).

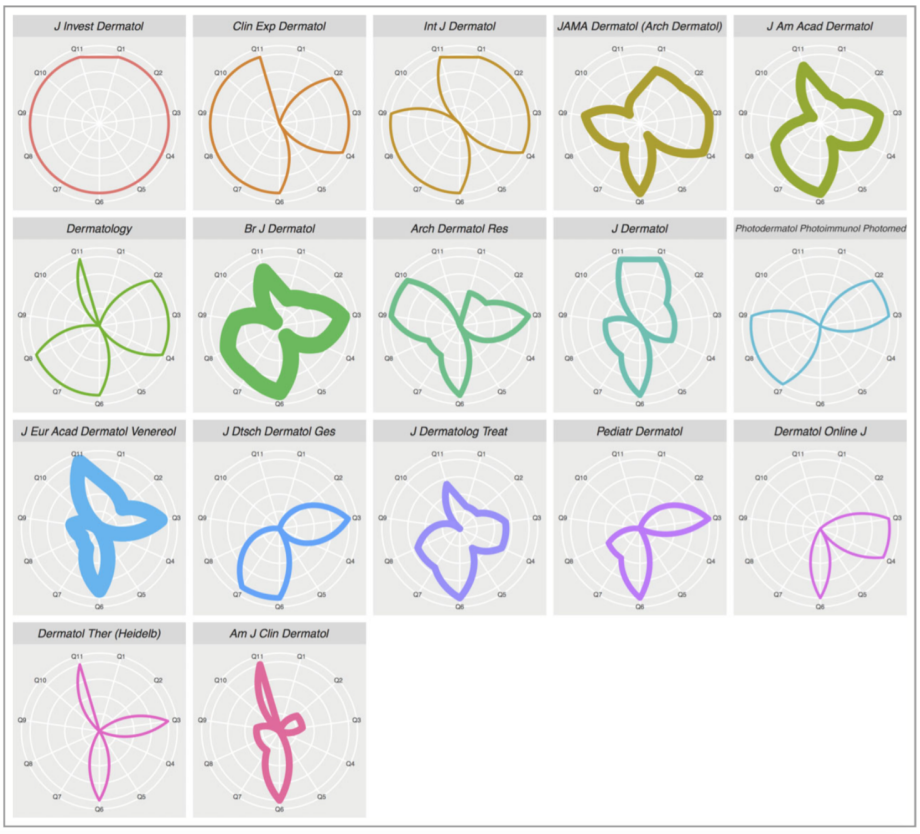


Fig. 3 Method quality profiles of systematic reviews for therapeutic interventions in psoriasis published in dermatology journals. A series of radial plots displays in polar coordinates the proportion of compliance with each assessed AMSTAR (Assessment of Multiple Systematic Reviews) item considering all reviews published per journal. Line size is proportional to the total number of reviews assessed per journal.

Clustering algorithms identified six distinct clusters of these differences (Fig. 4e). Articles that belonged to clusters 1, 4 and 5 had a lower quality than those classified in clusters 2, 3 and 6 (Fig. 4f). Cluster 5 almost exclusively contained low-quality reviews of RCTs. Cluster 1 is defined by higher numbers of authors and institutions per review, by the higher number of authors with conflicts of interest, and by the higher frequency of studies funded by pharmaceutical companies. In cluster 6 we found the longest reviews, of highest quality, funded by academic institutions, with a low number of authors with a conflict of interest, and focused on the analysis of RCTs of psoriasis treatments. All Cochrane reviews belonged to this cluster. Most reviews that included an MA of RCTs and observational studies were found in cluster 3. Most reviews in clusters 4 and 2 were of moderate methodological quality. Cluster 4 contains reviews of psoriasis treatments or comorbidities that were funded mainly by pharmaceutical companies, although some were funded by academic institutions. Most reviews in cluster 2 included an MA, were performed by a low number of authors with a conflict of interest, and had no source of funding communicated.

Quality of reviews and financial disclosures

We further divided reviews into subgroups according to several characteristics defined by extracted articles and journal metadata, and compared their methodological qualities using AMSTAR per item and total scores. When considering the funding source, reviews funded by academic institutions had 5.4-fold better methodological quality than the others, whereas reviews funded by pharmaceutical companies, or whose authors declared a conflict of interest, had a 2.0-fold or 1.2-fold odds, respectively, of having a lower AMSTAR level (Table 2). Reviews funded by academic institutions had AMSTAR scores above the mean, were classified as moderate or high quality, and had fewer authors with a conflict of interest (Fig. S1; see Supporting Information).

Table 8 Univariate and multivariate ordinal logistic regression analysis for association between Assessment of Multiple Systematic Reviews (AMSTAR) score-based levels as the dependent variable and article, journal and author metadata

	Model 1		Model 2	
Variable	OR (95% CI)	P-value	OR (95% CI)	P-value
Research field				
Treatment of psoriasis	1.18 (0.68-2.04)	0.56		
Psoriasis pathogeny	0.70 (0.34-1.43)	0.32		
Psoriasis comorbidities	0.74 (0.38-1.41)	0.36		
Economic evaluation	0.78 (0.38-1.41)	0.45		
Article metadata				

Page count	1.03 (1.01-1.05)	0.004	1.08 (1.02-1.15)	0.021
Number of authors	1.00 (0.94-1.07)	0.93		
Number of institutions	1.04 (0.95-1.14)	0.44		
Number of countries	1.27 (0.71-2.42)	0.44		
Cites on Google Scholar	1.00 (1.00-1.01)	0.28		
Cites on Web of Science	1.00 (0.99-1-02)	0.51		
Cochrane Collaboration affiliation Authors with conflict of interest	10.56 ($6.3x10^{-30}$-$2.36x10^{-38}$)		0.79	
Authors with conflict of interest	0.94 (0.89-0.99)	0.029	0.90 (0.82-0.99)	0.024
Funding by pharmaceutical industry	0.50 (0.28-0.88)	0.016		
Funding by academic institutions	5.37 (2.80-10-68)	0.001	2.90 (1.11-7.89)	0.032
RCTs as primary studies	2.03 (1.20-3.50)	0.009		
Observational studies as primary studies	1.20 (0.70-2.06)	0.50		
Systematic reviews as primary studies	0.64 (0.18-2.33)	0.49		
Economic analyses as primary studies	1.73 (0.23-13.35)	0.60		
Meta-analysis included	7.66 (4.08-15.20)	0.001	6.22 (2.78-14.86)	0.001
Economic analysis included	1.00 (0.012-80.85)	1.00		
'Dermatology' journal area	1.03 (0.61-1.73)	0.92		
Journal bibliometrics				
Journal impact factor	1.28 (1.09-1.55)	0.006		
Impact factor without journal self-cites	1.26 (1.07-1.54)	0.016		
Five-year impact factor	1.27 (1.07-1.57)	0.018	1.34 (1.02-1.40)	0.006
Immediacy index	1.76 (1.21-2.68)	0.004		
Cited half-life	1.02 (1.00-1.04)	0.071		
Citing half-Life	0.96 (0.92-0.99)	0.021		
Eigenfactor Score	1.66 (0.44-6.53)	0.45		
Normalized Eigenfactor	0.99 (0.98-1.01)	0.52		
Article Influence score	2.44 (1.34-4.67)	0.005	2.14 (1.05-6.67)	0.001
SJR	1.79 (1.26-2.59)	0.001		
SJR best quartile (reference Q1)				
Q2	0.48 (0.21-1.06)	0.067		
Q3	0.43 (0.11-1-69)	0.22		
Q4	0.50 (0.05-4.54)	0.52		
Journal H-index	1.01 (1.00-1.02)	0.001		

Author bibliometrics		
Total documents	1.00 (1.00-1.00)	0.001
Total citations	1.00 (1.00-1.00)	0.001
Total coauthors	0.99 (0.99-1.00)	0.001
Author H-index	0.98 (0.97-0.98)	0.001

Figures S1–S4 display time-course changes of total AMSTAR scores by the funding source and number of authors with a conflict of interest. The methodological quality improved from the first study published in 1997 to the final study published in 2016. Reviews funded by academic institutions had the highest-quality scores without significant variations. Conversely, studies funded by pharmaceutical companies or with an unknown funding source had the lowest AMSTAR scores, with a slight improvement in the most recent year of study for the former. Finally, reviews with authors who declared no funding source had an intermediate profile; they had a quality similar to those funded by pharmaceutical companies, showed progressive quality improvement from 2009 to 2016, and achieved AMSTAR scores close to those in articles funded by academic institutions.

Methodological quality and bibliometric indices

When analysing the quality of reviews by grouping the articles based on article metadata, some journal bibliometric indices showed a 1.2–2.4-fold probability that any published review had a higher methodological quality, including: Article Influence score (OR 2.44, 95% CI 1.34–4.67); SJR (OR 1. 79, 95% CI 1.26–2.59); immediacy index (OR 1.76, 95% CI 1.21– 2.68); 5-year impact factor (OR 1.27, 95% CI 1.07–1.57); journal impact factor (OR 1.28, 95% CI 1.09–1.55); and impact factor without journal self-cites (OR 1.26, 95% CI 1.07–1.54) (Table 2). Only citing half-life (OR 0.96, 95% CI 0. 92–0.99) was associated with lower AMSTAR quality levels.

Predictive model of methodological quality

Factors such as: including MA (OR 6.22, 95% CI 2.78– 14.86); funding by academic institutions (OR 2. 90, 95% CI 1.11–7.89); Article Influence score (OR 2.14, 95% CI 1.05–6.67); 5-year impact factor (OR 1.34, 95% CI 1.02–1.40); article page count (OR 1.08, 95% CI 1.02–1.15); and number of authors with a conflict of interest (OR 0.90, 95% CI 0.82–0.99), significantly predicted a higher quality in the final model (Fig. 5, Table 2). Results of k-fold cross-validation demonstrated that our model performed better in predicting low- vs. moderate- or high-quality reviews (j = 0 32, sensitivity 0.83, specificity 0.76, positive predictive value 0.50, negative predictive value 0. 94).

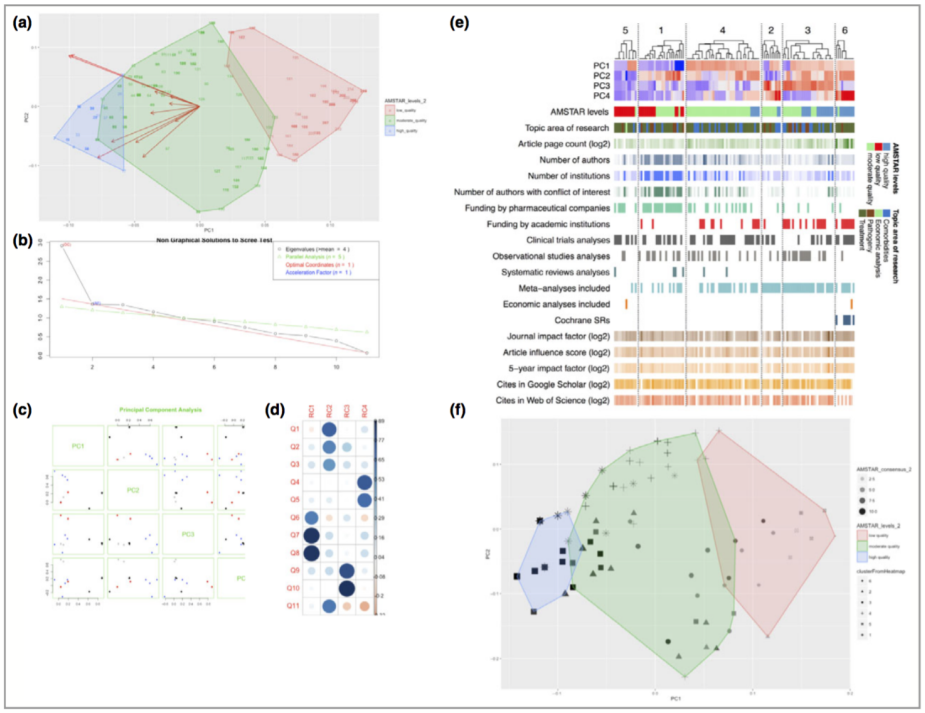


Fig. 4 Scale reduction of Assessment of Multiple Systematic Reviews (AMSTAR) items by principal component analysis (PCA) and article- and journal-related metadata clustering heat map. (a) Scatter plot displaying in two dimensions the relative position of every article based on the principal component (PC)1 and PC2 values obtained after PCA was performed using 11 AMSTAR items per article. (b) Graph showing the scree test results for PCA. (c) Matrix of scatter plots of pair comparisons for PC1–PC4. (d) Correlation matrix representing the relevance of each AMSTAR question for the first four component factors derived from PCA. (e) Clustering heat map of all include articles based on PC1–PC4 values per review. Three clusters (1–6) were indentified. Article- and journal-related metadata are also displayed as individual heat maps. (f) Scatter plot of AMSTAR items by cluster.

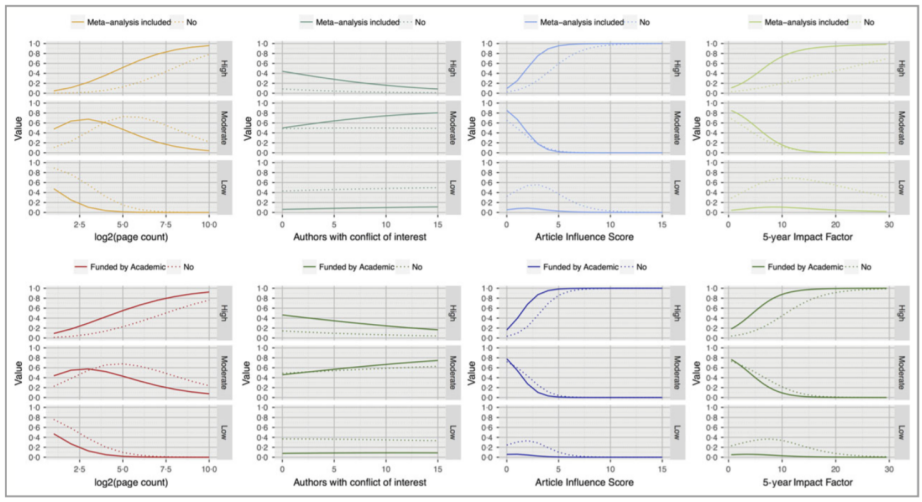


Fig. 5 Probability density plots. This panel displays the probability distribution of a review being classified as of high, moderate or low methodological quality using the estimates obtained in our regression model when meta-analysis is included and academic funding is declared, or when page count, number of authors with conflicts of interest, Article Influence score or 5-year impact factor are known.

Discussion

This is the first study to evaluate factors influencing the methodological quality of SRs and MAs on psoriasis. Our results suggest two hypotheses. Firstly, the relationship of pharmaceutical companies with authors (through research funding or author disclosures) may compromise the methodological quality. Secondly, reviews with poorer quality tend to be published in journals with lower bibliometric indices, although author-related bibliometric parameters do not influence the quality of these works. Therefore, there are publications of high and low quality from researchers with a high H-index. This means that a subgroup of researchers with high bibliometric scores may be involved in coauthoring reviews of low or moderate quality.

Industry funding accounts for more than half of biomedical research funding, and it is increasing in proportion to other funding sources.[11] In previous studies, there were associations between the author's or funder's conflict of interest and study outcomes,[12,13,14,15] between pharmaceutical industry relationship and author behaviour[16]or expressed opinions,[17] and between the reviewer's conflict of interest and conclusions in SRs.[18] In our study, we were able to predict that a higher number of authors with a conflict of interest would indicate lower methodological quality of an SR. Furthermore, we observed that research funded by academic institutions seemed to predict higher methodological quality of SRs as opposed to those studies funded by the pharmaceutical industry.

It is generally accepted that studies published in journals with a high impact factor are more important. The authors of these publications, in many cases from the academic arena, need more funding to continue their research. Among most governmental agencies or nonprofit institutions that finance research projects, evaluators positively value research that is published in journals with a high impact factor as a real reflection of the significance of the quality of research.[19,20] This means that they positively value papers published in journals that have accumulated a high number of citations over the previous years, and that they do not consider whether the work contained in their curriculum has really been highly cited by other researchers.

Fleming et al. analysed the overall percentage AMSTAR scores for each of 327 interventional SRs over a 6-month period during 2012, and they found that the methodological quality of reviews was better in journals with a higher impact factor.[21] Previous studies have shown similar results, and reporting details of RCTs tends to be more comprehensive in higher-impact journals.[22] Our data agree with these observations, but they also point to another issue: a critical appraisal of SRs published in lower-impact-factor journals is particularly important, but a high degree of suspicion should remain even for reviews published in journals with a higher impact factor,especially if they have other predictors of low quality (e.g. a high number of authors with a conflict of interest or no funding from academic institutions).

The pharmaceutical industry has considered that researchers can produce research more easily if they are provided with funding to achieve their goals. A working hypothesis suggests that the industry would act as an altruistic organization that funds projects and shares social commitments and research objectives with governmental or academic institutions to advance the science, without considering objectives related to the market sales of its products. But acceptance of this hypothesis implies that one would expect a similar quality of results, and as we have shown this is not so. Again, the only external factor is the participation of industry in the development and/ or publication of such works. The objective may be simple: to influence the content of the only articles most physicians have time to read (i.e. documents that synthesize evidence), which involves weighing the monetary resources available to patients with clinical decisions.

This may seem like a novel idea, but it is not. Several publications have previously shown the influence of conflicts of interest on the authors of RCTs,[15]or on those physicians who form the panels of experts who develop clinical practice guidelines.[23,24,25,26] These guidelines are closer to the decision-making processes that synthesize the evidence and establish recommendations based on the magnitude and quality of such evidence. Therefore, it would not be surprising that in previous stages of SRs, variations in the

methodology used may compromise the results and presented recommendations.[27] This article has many strengths. Our study includes the first large sample (n = 220) of SRs and MAs on skin psoriasis. We used an a priori protocol published in PROSPERO and sampled > 15 years of studies to evaluate trends in the literature. The AMSTAR score was determined independently by two authors, and there were few disagreements, all of which were solved by discussion.

The AMSTAR tool has been used previously in > 170 methodological quality-assessment articles in other fields of research, including cardiology, [28,29] gastroenterology, [30] gynaecology, [31,32,33,34] neumology,[35,36] neurology, [37,38,39] neurosurgery, [40,41,42] general surgery, [43,44] urology, [45,46] radiology, [47] and odontology.[48] Most studies achieved similar general conclusions to our study. However, AMSTAR is not a perfect tool for assessing methodological quality. One criticism of AMSTAR is that no guidance has been provided on how to translate total scores into categorical ratings. Various thresholds have been used to define categories for quality, making it difficult to compare assessments across reviews.[49] Thus, the validity of translating the AMSTAR scores into three categories (high, moderate and low methodological quality) is still unclear. Burda et al. recommended adding new items and modifying existing items to assess the quality of the body of evidence and to address subgroup and sensitivity analyses.[49] Faggion critically evaluated the ability of all AMSTAR checklist items to determine adequately the methodological quality of an SR, described difficulties regarding interpretation of the checklist, and provided potential solutions for these difficulties.[50]

One of the most debated aspects of AMSTAR is that it does not set different weights for each of the individual items that are evaluated.[50] This makes the contribution of each item to the total score the same, so there are articles with the same final value of AMSTAR from different items. Therefore, the differences between articles with the same final AMSTAR value can be determined only through a description of the discriminant items. PCA was used as an exploratory analytical tool to reveal the internal structure of our dataset to explain the variance in the quality of reviews better than simply by using the total score system. PCA successfully found linear combinations of different items that distinguished studies that had the same AMSTAR-based quality scores. However, some reviews were plotted in overlap regions between areas of high and moderate, and moderate and low quality. Interestingly, there appeared to be at least two clear subgroups in the moderate area that the classical AMSTAR three-level system did not capture. Comparing the capacity of AMSTAR vs. ROBIS to classify such revisions would be an interesting future project, given the prominence of ROBIS as a new rigorous tool. The ROBIS tool

was developed to employ accurate methodology across four wide groups of reviews within healthcare settings: aetiology, interventions, diagnosis and prognosis.[3]

Some studies have found that author affiliation to the Cochrane Collaboration is a predictor of the methodological quality of SRs and MAs.[51,52,53] Although this subgroup of reviews achieved the highest AMSTAR scores in our study, we did not observe Cochrane affiliation as a quality predictor. This may be due to the fact that most of these studies performed a linear regression using the total AMSTAR score as a predicting variable. The scarce number of Cochrane SRs of psoriasis found in our search and the low proportion of total SRs of high methodological quality in our dataset may be two additional factors that can explain this discrepancy. In any case, all Cochrane reviews were found in the cluster of studies with the highest AMSTAR score when PCA was performed.

Another potential limitation of our study is that, although the AMSTAR tool is widely used to evaluate the scientific quality of SRs, it has not been validated for SRs of nonrandomized studies. However, Pieper et al. found good psychometric properties, which were comparable with prior findings in SRs of RCTs, when assessing[32] SRs of nonrandomized studies investigating the hospital volume–outcome relationship in surgery.[7]

Finally, although a great amount of data were obtained following well-established methodology, the use of more useful tools such as decision trees or the development of a revised version of AMSTAR with weighted questions in the future may allow additional meaningful studies. Until further clarification, it would be inadvisable for readers to assess systematically the methodological quality of reviews without understanding the power and limitations of the tools used for assessment. Internal and external validation assessments are needed before applying our methods to other diseases (i.e. atopic dermatitis or rare diseases less influenced by pharmaceutical industry research).

In conclusion, the number of reviews published on psoriasis has increased substantially over time, but the methodological quality remains suboptimal. The fact that only 17% of reviews were of high methodological quality displays a bleak picture for evidence-based medicine in psoriasis. Some factors such the types of funding sources and author disclosures may compromise study quality, increasing the risk of methodological bias of SRs and MAs performed on this topic.

Acknowledgments

The authors would like to thank Editage (www.editage.com) for English-language editing.

References

[1] Goff KL, Karimkhani C, Boyers LN et al. The global burden of pso- riatic skin disease. Br J Dermatol 2015; 172:1665–8.
[2] Delgado-Rodrıguez M. Systematic reviews of meta-analyses: applications and limitations. J Epidemiol Community Health 2006; 60:90–2.
[3] Whiting P, Savovic J, Higgins JPT et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016; 69:225–34.
[4] Shea BJ, Grimshaw JM, Wells G et al. Development of AMSTAR: a measurement tool to assess the methodological quality of system- atic reviews. BMC Med Res Methodol 2007; 7:10.
[5] Shea BJ, Bouter LM, Peterson J et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One 2007; 2:e1350.
[6] Shea BJ, Hamel C, Wells GA et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009; 62:1013–20.
[7] Pieper D, Mathes T, Eikermann M. Can AMSTAR also be applied to systematic reviews of non-randomized studies? BMC Res Notes 2014; 7:609.
[8] Kelly SE, Moher D, Clifford TJ. Quality of conduct and reporting in rapid reviews: an exploration of compliance with PRISMA and AMSTAR guidelines. Syst Rev 2016; 5:79.
[9] Maneiro JR, Salgado E, Gomez-Reino JJ. Immunogenicity of mon- oclonal antibodies against tumor necrosis factor used in chronic immune-mediated inflammatory conditions: systematic review and meta-analysis. JAMA Intern Med 2013; 173:1416–28.
[10] Schlager JG, Rosumeck S, Werner RN et al. Topical treatments for scalp psoriasis. Cochrane Database Syst Rev 2016; 2:CD009687.
[11] Moses H, Matheson DHM, Cairns-Smith S et al. The anatomy of medical research: US and international comparisons. JAMA 2015; 313:174–89.
[12] Bekelman JE, Li Y, Gross CP. Scope and impact of financial con- flicts of interest in biomedical research: a systematic review. JAMA 2003; 289:454–65.
[13] Lexchin J, Bero LA, Djulbegovic B et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003; 326:1167–70.
[14] Stelfox HT, Chua G, O'Rourke K et al. Conflict of interest in the debate over calcium-channel antagonists. N Engl J Med 1998; 338:101–6.
[15] Als-Nielsen B, Chen W, Gluud C et al. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? JAMA 2003; 290:921–8.

[16] Chren MM, Landefeld CS. Physicians' behavior and their interactions with drug companies. A controlled study of physicians who requested additions to a hospital drug formulary. JAMA 1994; 271:684–9.
[17] Wang AT, McCoy CP, Murad MH et al. Association between indus- try affiliation and position on cardiovascular risk with rosiglitazone: cross sectional systematic review. BMJ 2010; 340:c1344.
[18] Dunn AG, Arachi D, Hudgins J et al. Financial conflicts of interest and conclusions about neuraminidase inhibitors for influenza: an analysis of systematic reviews. Ann Intern Med 2014; 161:513–18.
[19] Carey RM. Quantifying scientific merit. Circ Res 2016; 119:1273–5.
[20] van Eck NJ, Waltman L, van Raan AFJ et al. Citation analysis may severely underestimate the impact of clinical research as compared to basic research. PLoS One 2013; 8:e62395.
[21] Fleming PS, Koletsi D, Seehra J et al. Systematic reviews published in higher impact clinical journals were of higher quality. J Clin Epidemiol 2014; 67:754–9.
[22] Bala MM, Akl EA, Sun X et al. Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. J Clin Epidemiol 2013; 66:286–95.
[23] Choudhry NK, Stelfox HT, Detsky AS. Relationships between authors of clinical practice guidelines and the pharmaceutical industry. JAMA 2002; 287:612–17.
[24] Norris SL, Holmer HK, Ogden LA et al. Conflict of interest disclosures for clinical practice guidelines in the National Guideline Clearinghouse. PLoS One 2012; 7:e47343.
[25] Papanikolaou GN, Baltogianni MS, Contopoulos-Ioannidis DG et al.Reporting of conflicts of interest in guidelines of preventive and therapeutic interventions. BMC Med Res Methodol 2001; 1:3.
[26] Neuman J, Korenstein D, Ross JS et al. Prevalence of financial con- flicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. BMJ 2011; 343:d5621.
[27] Mandrioli D, Kearns CE, Bero LA. Relationship between research outcomes and risk of bias, study sponsorship, and author financial conflicts of interest in reviews of the effects of artificially sweetened beverages on weight outcomes: a systematic review of reviews. PLoS One 2016; 11:e0162198.
[28] KitsiouS,Par eG,Jaana M.Effects of home telemonitoring interventions on patients with chronic heart failure: an overview of systematic reviews. J Med Internet Res 2015; 17:e63.
[29] Anderson L, Taylor RS. Cardiac rehabilitation for people with heart disease: an overview of Cochrane systematic reviews. Cochrane Database Syst Rev 2014; 12:CD011273.

[30] Xin Y, Manson J, Govan L et al. Pharmacological regimens for eradication of Helicobacter pylori: an overview of systematic reviews and network meta-analysis. BMC Gastroenterol 2016; 16:80.
[31] Windsor B, Popovich I, Jordan V et al. Methodological quality of systematic reviews in subfertility: a comparison of Cochrane and non-Cochrane systematic reviews in assisted reproductive technologies. Hum Reprod 2012; 27:3460–6.
[32]Brown J, Farquhar C. Endometriosis: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2014; 3:CD009590.
[33]Hindocha A, Beere L, Dias S et al. Adhesion prevention agents for gynaecological surgery: an overview of Cochrane reviews. Cochrane Database Syst Rev 2015; 1:CD011254.
[34]Farquhar C, Rishworth JR, Brown J et al. Assisted reproductive technology: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2015; 7:CD010537.
[35]Reid WD, Yamabayashi C, Goodridge D et al. Exercise prescription for hospitalized people with chronic obstructive pulmonary disease and comorbidities: a synthesis of systematic reviews. Int J Chron Obstruct Pulmon Dis 2012; 7:297–320.
[36] Ho RST, Wu X, Yuan J et al. Methodological quality of meta-analyses on treatments for chronic obstructive pulmonary disease: a cross-sectional study using the AMSTAR (Assessing the Methodological Quality of Systematic Reviews) tool. NPJ Prim Care Respir Med 2015; 25:14102.
[37] Miyahara M. Meta review of systematic and meta analytic reviews on movement differences, effect of movement based interventions, and the underlying neural mechanisms in autism spectrum disorder. Front Integr Neurosci 2013; 7:16.
[38] Parke HL, Epiphaniou E, Pearce G et al. Self-management support interventions for stroke survivors: a systematic meta-review. PLoS One 2015; 10:e0131448.
[39] Hersi M, Quach P, Wang M-D et al. Systematic reviews of factors associated with the onset and progression of neurological conditions in humans: a methodological overview. Neurotoxicology 2016; pii: S0161-813X(16)30118-8.
[40] Klimo P, Thompson CJ, Ragel BT et al. Methodology and reporting of meta-analyses in the neurosurgical literature. J Neurosurg 2014; 120:796–810.
[41]Ding F, Jia Z, Zhao Z et al. Total disc replacement versus fusion for lumbar degenerative disc disease: a systematic review of overlap- ping meta-analyses. Eur Spine J 2017; 26:806–15.
[42]Santaguida PL, Keshavarz H, Carlesso LC et al. A description of the methodology used in an overview of reviews to evaluate evidence on the treatment, harms, diagnosis/classification, prognosis and outcomes used in the management of neck pain. Open Orthop J 2013; 7:461–72.

[43] Martel G, Duhaime S, Barkun JS et al. The quality of research syn- thesis in surgery: the case of laparoscopic surgery for colorectal cancer. Syst Rev 2012; 1:14.
[44] Potter S, Browning D, Savovi c J et al. Systematic review and critical appraisal of the impact of acellular dermal matrix use on the out- comes of implant-based breast reconstruction. Br J Surg 2015; 102:1010–25.
[45] Lee YJ, Park JE, Jeon BR et al. Is prostate-specific antigen effective for population screening of prostate cancer? A systematic review Ann Lab Med 2013; 33:233–41.
[46] Braga LH, Pemberton J, Demaria J et al. Methodological concerns and quality appraisal of contemporary systematic reviews and meta-analyses in pediatric urology. J Urol 2011; 186:266–71.
[47] Blomqvist L, Carlsson S, Gjertsson P et al. Limited evidence for the use of imaging to detect prostate cancer: a systematic review. Eur J Radiol 2014; 83:1601–6.
[48] Mathur S, Conway DI, Worlledge-Andrew H et al. Assessment and prevention of behavioural and social risk factors associated with oral cancer: protocol for a systematic review of clinical guidelines and systematic reviews to inform primary care dental professionals. Syst Rev 2015; 4:184.
[49] Burda BU, Holmer HK, Norris SL. Limitations of A Measurement Tool to Assess Systematic Reviews (AMSTAR) and suggestions for improvement. Syst Rev 2016; 5:1–10.
[50] Faggion CM. Critical appraisal of AMSTAR: challenges, limitations, and potential solutions from the perspective of an assessor. BMC Med Res Methodol 2015; 15:63.
[51] Fleming PS, Seehra J, Polychronopoulou A et al. Cochrane and non-Cochrane systematic reviews in leading orthodontic journals: a quality paradigm? Eur J Orthod 2013; 35:244–8.
[52] Wasiak J, Tyack Z, Ware R et al. Poor methodological quality and reporting standards of systematic reviews in burn care manage- ment. Int Wound J 2016; DOI: 10.1111/iwj.12692.
[53] Zhu Y, Fan L, Zhang H et al. Is the best evidence good enough: quality assessment and factor analysis of meta-analyses on depres- sion. PLoS One 2016; 11:e0157808.

Journal of Clinical Epidemiology

Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool

F. Gomez-Garcia[1,2,*], **J. Ruano**[1,2,*,&], **J. Gay-Mimbrera**[2], **M. Aguilar-Luque**[2], **J.L. Sanz-Cabanillas**[1,2], **P. Alcalde-Mellado**[2,3], **B. Maestre-Lopez**[2,3], **P. J. Carmona-Fernandez**[2], **M. Gonzalez-Padilla**[1,2], **A. Velez Garcia-Nieto**[1,2], **Beatriz Isla-Tejera**[2,4]. [1]Department of Dermatology, Reina Sofía University Hospital, 14004 Córdoba, Spain.
[2]IMIBIC/Reina Sofía University Hospital/University of Cordoba, 14004 Córdoba, Spain.
[3] School of Medicine, University of Cordoba, 14004 Córdoba, Spain.
[4]Department of Pharmacy, Reina Sofía University Hospital, 14004 Córdoba, Spain.
[&]**Corresponding author**. Tel.: +34 9 57011213; fax: +34 9 57011214.
E-mail: juanruanoruiz@mac.com
***Equal contributors.**

Funding sources: This work was supported, in part, by project ICI1400136 to JR, integrated into the National Plan of R+D+I 2008-2011 and co-financed by the ISCIII-Subdirección General de Evaluación and European Regional Development Fund (ERDF), and by grant PP13/009 of Plan Propio de movilidad para investigadores del Instituto Maimonides de Investigacion Biomédica de Córdoba (IMIBIC). No funding was received from any pharmaceutical company.

Conflict of Interest: FG-G has received honoraria for research from Pfizer, and for lecturing from AbbVie, Janssen-Cilag and Novartis; JR has received honoraria for lecturing and grants for research from Pfizer, honoraria for lecturing from Janssen-Cilag and Novartis, and other financial benefits from AbbVie and Novartis; AVG-N has received honoraria for lecturing from Pfizer, Novartis, AbbVie, and Janssen-Cilag, and other financial benefits from AbbVie, Novartis, and Janssen-Cilag. MA-L, JR, FG-G, and BI-T are members of the Cochrane Bias Methods Group and Skin Group. JLS-C, PA-M, BM-L, PJC-F, and MG-P have no disclosures.

Gómez-García F, Ruano J, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, Maestre-López B, Carmona-Fernández PJ, González-Padilla M, García-Nieto AV, Isla-Tejera B. Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool. J Clin Epidemiol. 2017 Sep 9. pii: S0895-4356(17)30534-6. doi: 10.1016/j.jclinepi.2017.08.015. [Epub ahead of print]

Abstract

Objectives: No gold standard exists to assess methodological quality of systematic reviews (SRs). Although Assessing the Methodolofical Quality of Systematic Reviews (AMSTAR) is widely accepted for analysing quality, the ROBIS instrument has recently been developed. This study aimed to compare the capacity of both instruments to capture the quality of SRs concerning psoriasis interventions.

Study Design and Setting: Systematic literature searches were undertaken on relevant databases. For each review, methodological quality and bias risk were evaluated using the AMSTAR and ROBIS tools. Descriptive and principal component analyses were conducted to describe similarities and discrepancies between both assessment tools.

Results: We classified 139 intervention SRs as displaying high/moderate/low methodological quality, and as high/low risk of bias. A high risk of bias was detected for most SRs classified as displaying high or moderate methodological quality by AMSTAR. When comparing ROBIS result profiles, responses to domain 4 signalling questions showed the greatest differences between bias risk assessments, while domain 2 items showed the least.

Conclusion: When considering SRs published about psoriasis, methodological quality remains suboptimal, and the risk of bias is elevated, even for SRs exhibiting high methodological quality. Further, the AMSTAR and ROBIS tools may be considered as complementary when conducting quality assessment of SRs.

Keywords: Psoriasis; Systematic Reviews; Risk of Bias; Methodological Quality; AMSTAR; ROBIS.

Introduction

Psoriasis is a chronic disease, with moderate and severe forms associated with significant comorbidity, impaired quality of life, and high direct and indirect costs.[1] New therapies have been developed during the last decade that have been increasingly effective, but with potentially significant adverse side effects and higher costs, which puts patients at risk and calls into question the sustainability of health systems.[2-3] Therefore, therapeutic decision-making processes about appropriate psoriasis interventions should be based on the best evidence.[4]

Systematic reviews (SRs) are the standard for the synthesis of the evidence. Their conclusions are often used as a starting point for the development of clinical practice guidelines, establishing the recommendations of diagnostic, prognostic, and/or therapeutic interventions.[5] Making decisions based on SRs can improve health outcomes.[6] However,some research groups have found discrepancies in conclusions presented by SRs performed to response the same research question.[7] This type of disparity is a risk factor for health authorities and clinicians,which ultimately affects preferences of patients for specific interventions. This risk is even greater now,given that there has been a proliferation in recent years of incompetently conducted SRs, and SRs not intended to improve evidence, but rather, to use the prestige of journals to convey self-serving information.[8]

Multiple quality assessment tools have been developed to assess the methodological quality of reviews, although no one single tool has been universally accepted.[9,10] The most commonly used instrument is the measurement tool Assessing the Methodological Quality of Systematic Reviews (AMSTAR), an 11-items checklist that it has been proven to be both reliable and valid.[11,12] AMSTAR items can be used individually (components) or as a checklist by summing item scores into an overall score. However, this tool is somewhat limited in its ability to capture the overall quality of SRs. On one hand, the importance of each component varies, depending on the nature of the review; on the other, if even the total score is significant, this fact does not ensure a relationship between methodological quality and the risk of SR bias. [13]

Our group has previously used AMSTAR to evaluate the methodological quality of studies about psoriasis.[14] In this sense, Gómez-García et al. demonstrated that higher quality reviews, including meta-analyses (MAs), were funded by academic institutions, had fewer authors, and had a high article influence score. Reviews that contained a high number of authors with conflicts of interest were of lower quality. Sanz-Cabanillas et al. have found that structural differences in author-paper affiliation network may influence the methodological quality of these reviews, and authors who maintain an appropriate

balance between scientific quality and productivity are more likely to develop higher quality reviews.[15]

Because of the previously mentioned information, it is a priority to identify SRs of higher quality before making decisions. Because the objective of conducting SRs is to minimize bias, the measurement of methodological quality should be linked to the risk of bias. In this sense, Moher et al. defined the quality of SRs as the probability that the design will generate non- skewed results.[16] The ROBIS tool has recently been developed with the intention to measure the bias risk of SRs.[17,18] This tool is based on the evaluation of domains of items and how the methodological limitations are taken into account by the authors when drafting conclusions. Finally, while the tool presents a level of bias risk as high or low, it does not consider conflicts of interest or funding that may influence the quality and/or direction of the findings, which has been previously demonstrated. [19-21]

The aim of this study was to describe the relationship between AMSTAR components and ROBIS domains used to capture the methodological quality and bias risk of SRs concerning interventions in psoriasis.

Methods

2.1. Protocol and elegibility criteria

We established an a priori protocol and published it in the PROSPERO International Prospective Register of Systematic Reviews (PROSPERO 2016: CRD42016053181). We included SRs or MAs published concerning interventions of skin psoriasis in scientific journals.Abstracts of congresses, case reports, surveys, narrative reviews, narrative reports (i.e. those with a focus on understanding a concept), clinical practice guidelines, consensus documents, or MAs performed without a systematic literature search, and reviews titled as literature reviews or integrative reviews were not included. Our retrieval was restricted to English language reviews because of time limitations for project completion. There was no limitation on the year of publication or study population.

2.2. Search and selection methods

We filtered results obtained in a previously systematic literature search up to 5 July 2016 and published in a previous study.[14] New SRs and MAs published up to January 2017 were identified in MEDLINE, EMBASE, and the Cochrane Database. Details regarding the search methods for identifying and selecting are provided in the Supplementary materials

and methods (Supporting Information).Lists of included and excluded studies are shown in Tables S1 and S2.

2.3 Evaluation with AMSTAR instrument

Two investigators (F.G.-G., and J.G.-M.) independently assessed the methodological quality of each review blinded to the name of the journal, the name of the authors or the affiliations, and using the same data abstraction forms and 11-point AMSTAR criteria. In case of disagreement an independent researcher (J.R.) was consulted. Detailed information for the AMSTAR checklist and the system of rating the articles are presented in Supplementary materials and methods and Table S3 of supplementary information.

2.4. Assessment using ROBIS tool

Two investigators (F.G.-G., and M.A.-L.) independently assessed the risk of bias of each review using the same data abstraction forms and blinded to the name of the journal, the name of the authors or the affiliations. We used ROBIS, which is a four-stage approach to assess the bias risk of systematic reviews of psoriasis interventions. [9] ROBIS is completed in 3 phases. Phase 1 assesses the relevance the review, and is considered optional. Phase 2 includes four domains that covers (1) study eligibility criteria, (2) identification and selection of studies, (3) data collection and study appraisal, and (4) synthesis and findings. Phase 3 assesses the overall risk of bias in the interpretation of review findings and whether limitations identified in any of the phase 2 domains are considered. Phase 1 assesses the relevance of the review and is optional.

2.5. Data extraction and statistical analysis

For studies that fulfilled the inclusion criteria,five investigators (F.G.-G., J.G.-M., P.A.-M., J.L.S.-C, and M.G.-P.) independently obtained metadata from every article. Studies were classified as Cochrane vs non Cochrane reviews based on the authors' affiliation to the Cochrane group. Principal component analysis (PCA) was performed to discover potentially significant subgroups beyond the classical AMSTAR three-level classification. AMSTAR total scores are summarized descriptively as a median and interquartile range. AMSTAR and ROBIS results are also summarized as a percentage of achievement per item. A correlation matrix of ROBIS items was obtained. We used RadViz, a projection-based multivariate visualization R package, to arrange signalling questions of ROBIS in radial layouts. Although a PCA scatter plot displays reviews by using PCA and PC2, a RadViz

plot projects the nonlinearly normalized responses to all signalling questions for each review. In the case of RadViz, the influence of each question can be interpreted as a balance between the influences of all questions. Question order was optimized based on the cosine distance between questions, so that highly correlated questions were placed close together on the circle. Statistical significance was set at $P > 0.05$.

Graphs were produced and statistics were analysed using several packages of the R language (R Development Core Team). Our analysis can be fully reproduced by using several source files containing raw data and R scripts stored at our GitHub hosting repository.

2.6. Protocol vs. overview

Our planned search strategy recorded in PROSPERO was compared with the final reported review methods. However, our retrieval was restricted to English language reviews because of time limitations for project completion. We did not add, omit, or change outcomes after our protocol was published.

2.7. Ethical considerations

Because our study did not collect primary data, no formal ethical assessment and informed consent were required.

Results

3.1. Search results

Our new database search (from 5 July 2016 to 01 January 2017) yielded 161 titles with potential relevance (125 EMBASE & MEDLINE, 10 EMBASE only, 3 MEDLINE only, and 23 Cochrane Database). After excluding duplicated articles and screening title and abstracts, 44 new studies were judged potentially eligible for full-text review that were summed to the previously obtained 119 reviews (Fig. S1).

3.2. General characteristics

Thus, 139 reviews comprising 4,357 primary studies about interventions in psoriasis and published by 857 authors in 62 peer-reviewed journals were assessed (Tables S1 S2). The median numbers of authors and primary studies per review were 5 (range 2-20) and 21

https://github.com/info4cure/ROBISvsAMSTARassessment

(range 1-312), respectively. Only nine (6.4%) reviews were undertaken by Cochrane affiliated authors.

3.3. Results using AMSTAR instrument

Assessment of the methodological quality using AMSTAR questions (Q) began after agreement among reviewers was substantial (k = 0.75; 95% CI, 0.69-0.82). The median AMSTAR score was 7 (interquartile range, 1-11). Reviews were classified as 271 displaying high (22.3%), moderate (53.2%), or low (24.5%) methodological quality. AMSTAR items with the lowest compliance rates were Q5 (*'list of studies provided'*, 17.9%), Q10 (*'publication bias assessed'*, 20.1%), Q1 (*'a priori design provided'*, 46%), and Q4 (*'status of the publication included'*, 52.5%). Q6, *'which assessed whether characteristics of the included studies were provided'*, had the best compliance rate in the AMSTAR checklist (90.6%). Total AMSTAR scores achieved by Cochrane reviews were 10 or more.

3.4. Results using ROBIS tool

When the risk of bias was assessed using ROBIS, the percentage of rater agreement was lower than what was observed with AMSTAR rater agreement (k = 0.70; 95% CI, 0.66-0.81). Reviews were classified as high (86%) or low (14%) risk of bias. Phase 2 domains (D) of ROBIS with the highest amount of concern were D3 (*'data collection and study appraisal'*, 91%), and D1 (*'study eligibility criteria'*, 90%;Fig. 1a).ROBIS signalling questions (QR) with the highest concerns (i.e. higher rates of 'no' and 'probably no' responses) were QR45 ('*Were the findings robust as demonstrated through funnel plot or sensitivity analyses?*', no = 12.9%) and QR33 ('*Were all relevant study results collected for use in the synthesis?*', no = 20.1% and probably no = 66.9%; Fig. 1b). QR21, which assessed whether the *'search includes an appropriate range of databases/electronic sources for published and unpublished reports'*, displayed the best compliance rate (90%). All Cochrane reviews demonstrated a low risk of bias based on the ROBIS tool.

3.5. Multidimensional scaling

We used PCA to convert vectors of 11 AMSTAR item subscores and answers to the 21 ROBIS signaling questions per article into two sets of values of linearly uncorrelated variables called principal components (PCs), or projections to anchored domains or questions, respectively. Figs. 2A and D show two PCA scatterplots that comprise PC1 and PC2 projections of all included reviews. Overlapping was more evident between high or low

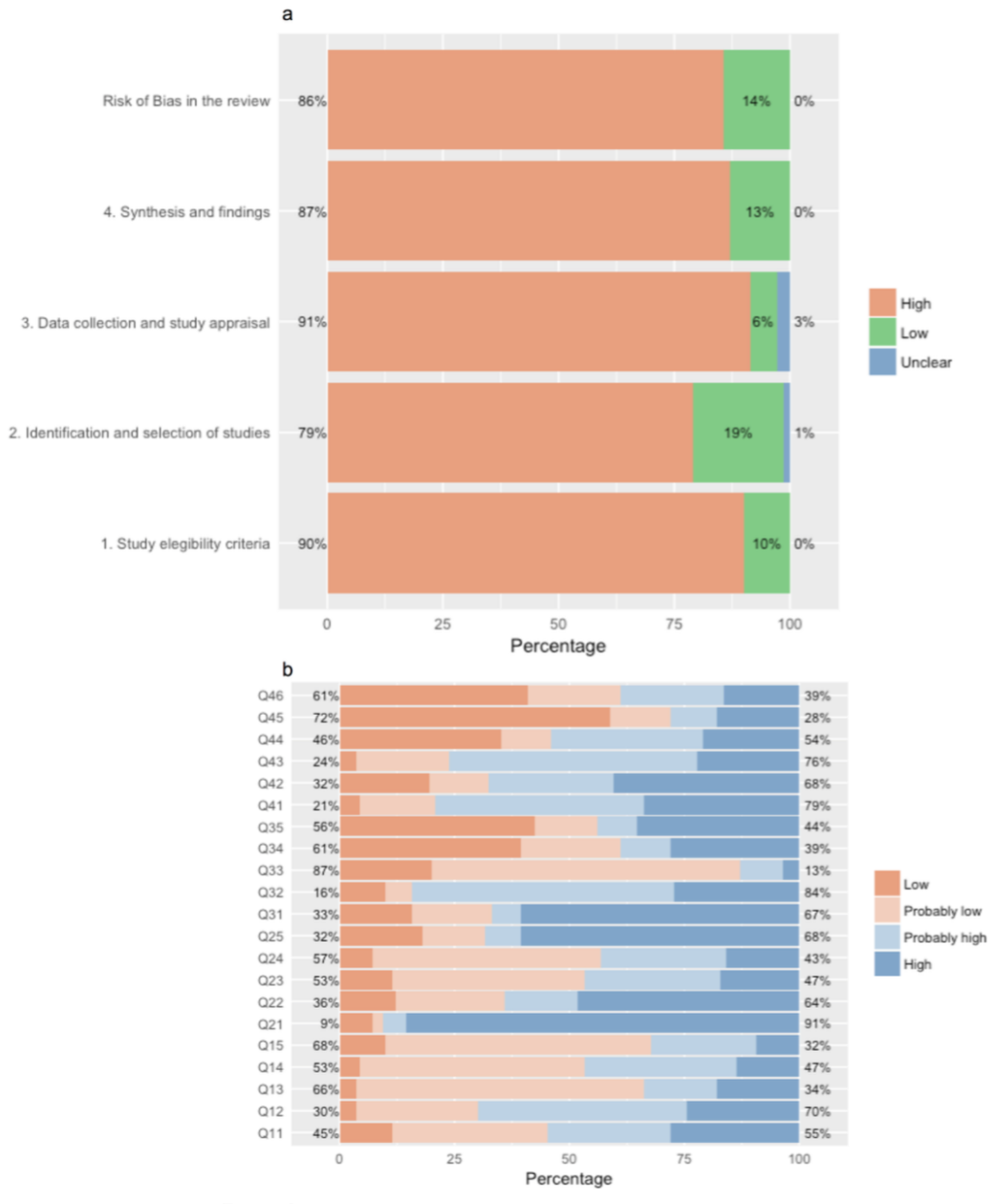


Fig. 1 Plot of Likert scales with ROBIS tool: overall risk of bias, dominion judgment, and response to signaling questions. This panel shows the frequency distributions of responses to risk of bias assessment using ROBIS tool. (a) This graph shows frequency distributions of potencial for concern of risk of bias by each Phase 2 domain ('high', 'low', or 'unclear') and overall risk of bias judgment by review ('high' vs 'low'). (b) This plot displays frequency distributions of responses ('low', 'probably low', 'probably high', or 'high') to signaling questions of Phase 2 domains.

Fig. 2 Scale reduction and high-dimensional visualization of AMSTAR and ROBIS results using principal component analysis (PCA). PCA was performed using the 11 AMSTAR item subscores or the responses to 21 signaling questions of ROBIS per article. Fig 2a and 2d display PC1-PC2 projections of every review, using different shapes to identify the risk of bias or the level of methodological quality. Fig. 2b and 2e show the contribution of each variable on PC1 and PC2. Fig. 2c and 2f display the contribution of each review to PC1 and PC2. A color gradient represent the magnitude of variables and review contributions. Fig. 2g-j show Radviz data visualization of ROBIS phase 2 domains judgments and response profiles to signaling questions. Points represent reviews and are colored with respect to the risk of bias (low-turquoise, high-coral) or methodological quality classification (high-blue, moderate-green, low-red).

risk of bias reviews compared with reviews demonstrating high vs. moderate, or moderate vs. low methodological quality. A scree plot of AMSTAR-based PCA data showed that the first component (PC1) explained 30% of variance, and that components PC1, PC2, PC3, and PC4 explained more than 50% of this variability (Fig. S2a). When considering ROBIS-based PCA data, the scree plot displayed a different result: component PC1 explained more than 45% of variance, while each of the following components contributed individually to less than 5-10% to this variability (Fig. S2b).

We further analysed how each item or question contributed to explain the observed variability . For AMSTAR-based PCA, QA8 (*'Was the scientific quality of the included studied used appropriately in formulating conclusions?'*) and QA7 (*'Was the scientific quality of the included studies assessed and documented?'*) were the items that contributed the most

to discriminate between reviews, while QA6 (*'Were the characteristics of the included studies provided?'*) contributed the least (Fig. 2c). In the case of ROBIS-based PCA, QR5 (*'Were efforts made to minimize error in risk of bias assessment?'*) and QR34 (*'Was risk of bias [or methodologic quality] formally assessed using appropriate criteria?'*) were the signaling questions that most contributed to explain variability of risk of bias among reviews (Fig. 2d).

Fig. 2g and h represents Radviz plots showing how the 21 signaling questions separate high and low risk of bias reviews apart. While low risk reviews clustered together in the center of the circle, high risk reviews were more sparced and some of them were overlapping with the formers. This fact shows that no perfect separation between low and high risk reviews is obtained when response to all signaling questions are considered. In Fig. 2i and j reviews are tagged by colors based on methodological classification by AMSTAR.Fig. S4 is a network plot displaying the relationship among ROBIS questions once those with a Spearman correlation coefficient > 0.5 were selected (Fig. S3). Nodes represent ROBIS questions and edges connect two questions if there is a significant correlation of results between them. The color and the width of each edge represent the magnitude of the correlation between connected nodes. QR34 (*'Was risk of bias formally assessed using appropriate criteria?'*), QR35 (*'Were efforts made to minimize error in risk of bias assessment?'*), and QR46 (*'Were biases in primary studies minimal or addressed in the synthesis?'*) represent a core hub externally connected with other nodes and focused on risk of bias questions. Interestingly, by displaying the network with nodes distributed in a plane based on PC1-PC2 coordinates, we observed that this hub is characterized by having the highest correlation coefficients and contributing the most to explain reviews variability (Fig. S5).

3.6. Relationships between AMSTAR and ROBIS results

Fig. 3a and Fig. 3b are Likert plots that display item subscores and question responses by grouping the reviews based on the risk of bias by the ROBIS tool or AMSTAR-based levels of methodological quality, respectively. A mosaic plot displays crossed frequencies of review classification by AMSTAR vs ROBIS tools (Fig. 3c), showing that more than 50% of SRs classified as high methodological quality were also at high risk of bias.

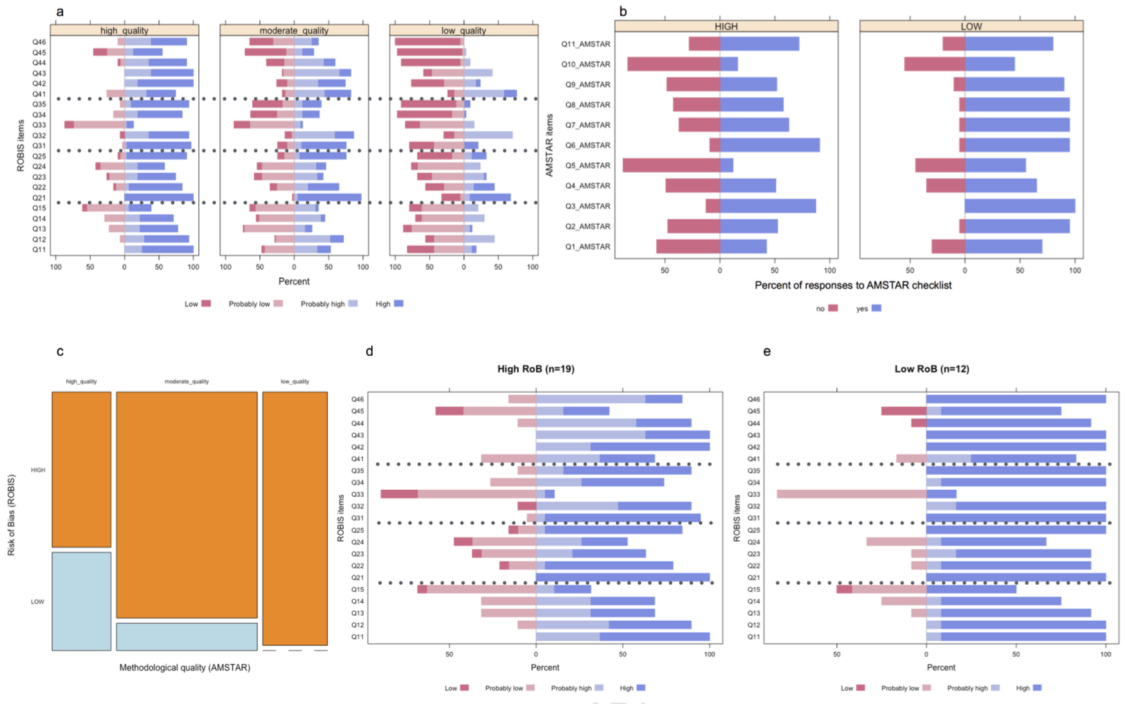


Fig. 3 Comparation of methodological quality vs risk of bias assessments. This panel of plots contains different graphs showing relationships between methodological quality and risk of bias when reviews are subgrouped by AMSTAR and ROBIS results. (a) This plot displays frequency distributions of responses ('low', 'probably low', 'probably high', or 'high') to signaling questions of ROBIS Phase 2 domain comparing reviews by AMSTAR-derived methodological quality levels ('high', 'moderate', or 'low'). (b) This plot shows frequency distributions responses ('no' or 'yes') of AMSTAR per item subscores comparing reviews by risk of bias using ROBIS tool ('high' or 'low'). (c) Mosaic plot that represent a contingency table comparing frequency distributions AMSTAR- derived methodological quality levels vs ROBIS-based risk of bias assessment. (d) and (e) graphs show frequency distributions of responses ('low', 'probably low', 'probably high', or 'high') to signaling questions of Phase 2 domain comparing reviews by overall risk of bias judgment using ROBIS tool ('high' or 'low') in a subset of reviews of high methodological quality based on AMSTAR instrument.

Fig. 3d and Fig. 3e show response profiles to signaling questions when reviews of high methodological quality were grouped based on the risk of bias (high vs low). QR33 was the question with least differences between both groups of reviews.

Discussion

4.1. Main findings

This is the first study that has compared results of the AMSTAR and ROBIS tools for examining methodological quality assessment, and it has highlighted areas of particular concern with regard the risk of bias in SRs concerning psoriasis interventions.

Overall, our results suggest that methodological quality only explains a proportion of the bias risk of SRs, as we observed that most of reviews classified as high and moderate methodological quality by AMSTAR were also considered as displaying a high risk of bias using ROBIS. The quality of SRs about psoriasis interventions is suboptimal, and that the risk of bias, including SRs of high quality, is elevated. Therefore, similar to the evaluation of primary studies, it is possible to carry out a SR following the highest methodological standards and still having a high risk of bias.[17] Two validated instrument for the assessment of methodological quality of SRs were used in this study. AMSTAR, a scale instrument that consists of 11-items annotated individually (components) or as the sum of reported items (overall score),[18] and ROBIS, a new tool developed as a domain-based approach and supported by signaling questions that follows the most recent risk assessment methods.[3] There are two fundamental differences of construct between AMSTAR and ROBIS. Although AMSTAR does not include in the assessment of how authors collected relevant data for the review, ROBIS does not take into account the compliance with the notification of authors' conflict of interests and/or funding sources.

Our data also demonstrate differences between both instruments when applied to the same set of reviews. It is known that the presence of missing data is a common problem in clinical trials. In a recent survey of five general medical journals, 87% in relation to the process of extracting data from relevant SR results. This fact raises problem in the conduct of the MAs and reduces the confidence in the estimates of the effects of SRs.[23] A recent study has also shown how the type of funding influences the methodological quality of SRs about psoriasis.[12] For instance, although industry-funded studies tend to be well-resourced,performed by highly skilled and experienced professionals, and based on detailed and extensively documented standarised procedures, we have demonstrated that many SRs of high methodological quality are still at high risk of bias. Although funding sources are related to the definition of bias, and there is enough empirically based evidence of bias related to this feature, neither the ROBIS instrument nor the Cochrane risk of bias tool include funding source as a standard item for risk of bias assessment of SRs or clinical trials respectively.[19]

We also have observed differences when 'inclusion of unpublished studies' item was evaluated using AMSTAR (39.5%) or ROBIS (64%). Previous studies using the AMSTAR tool have reported similar results (20-40.3%). [25,26] This fact corresponds to a difference criterion used to evaluate this item in each review. The important aspect here is that the evidence has found that SRs that exclude grey literature may lead to a hyperestimation of intervention effects.[27] This can be explained because the criterion of AMSTAR is more demanding than ROBIS.

In our study, signalling questions of the ROBIS instrument contributed differently to final discrimination between low and high bias risk in SRs. There are many reasons that make us to consider that QR33 signaling question (*'Were all relevant study results collected for use in the synthesis?'*) is a particularly useless signalling question. First, in our study, QR33 had low correlation coefficients with the rest of signaling questions. Second, the contribution of QR33 to PC1/PC2 in the ROBIS-related PCA was the lowest. Third, QR33 was the signaling question that discriminates the lowest between reviews when they were classified following AMSTAR-based methodological quality levels. Finally, QR33 does not vary significantly between high quality SRs demonstrating high or low-risk of bias. This is because the authors of low-risk bias SRs take into account in their considerations the implications of the missing data on the results. Therefore, it would be interesting to consider either simplifying or prioritizing the list of original signal questions included to reduce the time needed for the evaluation of each review. One of assessed items that showed low performance rating was 'publication bias', defined as the 'Achilles' heel' of SRs (AMSTAR, 27.8%; ROBIS, 28%). These frequencies are similar to the 21.8% found by Attkapo et al. that performed bias assessment of SRs and MAs published in 10 dermatology journals from 2006 to 2016 was performed.[24] Third domain aims to assess whether bias have been introduced.[9] We found that the third domain, and specifically QR34 and QR35, are the best to differentiate high vs low risk of bias among high methodological quality SRs. The aim of these signalling questions is to explore if the risk of bias of primary studies has been evaluated. Answers to these questions are fundamental to establish the validity of SRs results. An MA of biased effect estimates will likely produce a biased pooled analysis with increased precision and greater credibility.

4.2. Limitations and strengths

In this methodological study, we compared the ROBIS instrument and AMSTAR tool for assessing the quality of SRs, allowing to better establish the empirical difference between the concepts of 'methodological quality' and 'risk of bias'. Our study includes the first large sample of $>$ 15 years of reviews (n = 139) about interventions on psoriasis. In addition, the present study has been performed using a systematic search strategy and following an a priori protocol published in PROSPERO. The AMSTAR score was performed independently by two authors, and there were few disagreements, all of which were solved by discussion.

The search was restricted to MEDLINE, EMBASE, and Cochrane Databases, because our intention was to obtain a representative sample of published systematic reviews on psoriasis interventions, rather than cover all such reviews. We did not seek SRs in grey literature databases, and, therefore, we cannot establish differences of methodological

quality and risk of bias with respect to those that were examined.A limitation of this work is that we did not randomize the order in which the raters reviewed the articles or the order in which the evaluation was performed with both tools. Finally, only one of three raters carried out the evaluations both with AMSTAR and ROBIS tools. Although their results were compared in pairs and discrepancies were discussed with a fourth rater, there is a risk that this issue will affect the validity of our results.

4.3. Our findings in context

ROBIS is a new tool developed and validated for assessing the risk of bias in SRs, overcoming limitations of other instruments.[9] Although many studies have assessed the methodological quality of SRs using AMSTAR in a variety of research fields,[24-29] there are three SRs,[30-32] one overview of SRs,[33] and one umbrella review of MAs[34] that have used ROBIS tool to evaluate the risk of bias. These studies found that most reviews were rated as of high or unclear risk of bias across all ROBIS domains, except for those developed by Cochrane affiliated authors. Until the present moment, three full protocols have been recently published in scientific journals including the ROBIS tool for methodological quality assessment of SRs,[40-42] and one protocol has considered both AMSTAR and ROBIS to evaluate methodological quality of reviews.[43] Up to April 2017, there have been registered 54 protocols that have included the ROBIS tool and 18 records that included both ROBIS and AMSTAR instruments in the 'assessment of bias' field of the PROSPERO repository.[44]

Recently new tools have been validated to assess the risk of bias of SRs in specific fields of research. Faillie et al. have developed PROTECT, a tool specifically designed to evaluate the risk of bias in the context of drug safety assessment.[40] This tool includes eight domains: study design and objectives, selection bias, attrition, adverse events information bias, other information bias, statistical methods to control confounding, other statistical methods, and conflicts of interest. Interestingly, the total number of questions of this instrument varied from ten to 32 depending on the study design. So it can be considered as a flexible instrument that can be adapted to the type of study.

4.4. Implications of results

The AMSTAR tool has many limitations for assessing methodological quality of SRs. One criticism of AMSTAR is that no guidance has been provided on how to translate the total score into categorical ratings. Various thresholds have been used to define categories for quality, making it difficult to compare assessments across reviews.[46] Thus,the validity of

translating the AMSTAR scores into three categories (high, moderate, and low methodological quality) is still unclear. Some authors have recommended adding new items and modifying existing items to assess the quality of the body of evidence and to address subgroup and sensitivity analyses. The ability of all AMSTAR items to adequately determine the methodological quality of an SR has been questioned, considering difficulties regarding interpretation of the checklist, and providing potential solutions for these difficulties.[12]

One of the most debated aspects of AMSTAR is that it does not set different weights for each of the individual items that are evaluated.[40] This makes the contribution of each item to the total score the same, so there are articles with the same final AMSTAR value that are arguably very different methodologically. Therefore, the differences between articles with the same final AMSTAR value can only be determined through a description of the discriminant items. PCA was used as an exploratory analysis tool to reveal the internal structure of our dataset, in an effort to better explain the variance in the quality of reviews, rather than simply using the total score system. PCA successfully found linear combinations of different items that distinguished studies that had the same AMSTAR based quality scores.

5. Conclusions

The methodological quality of SRs published concerning psoriasis interventions remains suboptimal and the risk of bias is elevated even for most of the studies that demonstrated the highest levels of methodological quality. We recommend to use both AMSTAR and ROBIS tools when conducting quality assessment of SRs, as they may be considered as complementary instruments. An effort to simplify or stratify the list of signaling questions of ROBIS tool is desirable. Finally, given the small proportion of SRs with low risk of bias it would be advisable to implement the use of this type of instruments by editors, authors and/or reviewers to make clear the interpretation of their results.

Acknowledgments

The authors wish to acknowledge the assistance received from Dr. Penny Whiting, School of Social and Community Medicine, University of Bristol, Bristol, UK, for answering our questions about some aspects of the ROBIS tool during the evaluation phase of the reviews. The authors would like to thank Editage (www.editage.com) for English-language editing.

Supplementary data

References

[1] Goff KL, Karimkhani C, Boyers LN, Weinstock M, Lott JP, Hay RJ, et al. The Global Burden of Psoriatic Skin Disease. Br J Dermatol 2015;172:1665-8.
[2] Gómez-García F, Epstein D, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis. Br J Dermatol 2017;176:594-603.
[3] Nast A, Jacobs A, Rosumeck S, Werner RN. Efficacy and Safety of Systemic Long-Term Treatments for Moderate-to-Severe Psoriasis: A Systematic Review and Meta-Analysis. J Invest Dermatol 2015;135:26418.
[4] Dias S, Welton N. NICE DSU Technical Support Document 1: Introduction to evidence synthesis for decision making. Nice Dsu 2011;(April 2011):124.
[5] Abuabara K, Freeman EE, Dellavalle R. The role of systematic reviews and meta-analysis in dermatology. J Invest Dermatol 2012;132:e2.
[6] Lewis SJ, Orland BI. The Importance and Impact of Evidence Based medicine. J Manag Care Pharm 2004;10:S35.
[7] Low J, Ross JS, Ritchie JD, Gross CP, Lehman R, Lin H, et al. Comparison of two independent systematic reviews of trials of recombinant human bone morphogenetic protein-2 (rhBMP-2): the Yale Open Data Access Medtronic Project. Syst Rev 2017;6:28.
[8] Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. Milbank Quarterly 2016:485514.
[9] Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016;69:22534.
[10] Whiting P, Davies P, Savovic J, Caldwell D, Churchill R. (September 2013) Evidence to inform the development of ROBIS, a new tool to assess the risk of bias in systematic reviews, Available from http://www.robis-tool.info [accessed 11/08/2017]
[11] Shea BJ, Grimshaw JM, Wells G a, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.
[12] Faggion CM. Critical appraisal of AMSTAR: challenges, limitations, and potential solutions from the perspective of an assessor. BMC Med Res Methodol 2015;15:63.

[13] Oliveras I, Losilla JM, and Vives J. Methodological quality is underrated in systematic reviews and meta-analyses in health psychology. Journal of Clinical Epidemiology 2017;86:59-70.
[14] Gómez-García F, Ruano J, Aguilar-Luque M, Gay-Mimbrera J, Maestre-López B, Sanz-Cabanillas JL, et al. Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest, and bibliometric indices as predictors of methodological quality. Br J Dermatol 2017 Feb 13. [Epub ahead of print]
[15] Sanz-Cabanillas JL, Ruano J, Gómez-García F, Alcalde-Mellado P, Gay-Mimbrera J, Aguilar- Luque M, et al. Author-paper affiliation network architecture influences the methodological quality of systematic reviews and meta-analyses of psoriasis. PLoS One 2017;12:e0175419.
[16] Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials 1995;16:6273.
[17] Farquhar CM, Showell MG, Showell EAE, Beetham P, Baak N, Mourad S, et al. Clinical trial registration was not an indicator for low risk of bias. J Clin Epidemiol 2017.
[18] Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009;62(10):101320.
[19] Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. JAMA 2003;289:454-65.
[20] Lexchin J, Bero LA, Djulbegovic B et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003;326:1167-70.
[21] Stelfox HT, Chua G, ORourke K et al. Conflict of interest in the debate over calcium-channel antagonists. N Engl J Med 1998;338:101-6.
[22] Akl EA, Briel M, You JJ, Sun X, Johnston BC, Busse JW, et al. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. BMJ 2012;344:e2809e2809.
[23] Akl EA, Kahale LA, Agarwal A, Al-Matari N, Ebrahim S, Alexander PE, et al. Impact of missing participant data for dichotomous outcomes on pooled effect estimates in systematic reviews: a protocol for a methodological study. Syst Rev 2014;3:137.
[24] Barnes DE, Bero LA. Why review articles on the health effects of passive smoking reach different conclusions. JAMA 1998;279:156670.
[25] Liu D, Jin J, Tian J, Yang K. Quality assessment and factor analysis of systematic reviews and meta-analyses of endoscopic ultrasound diagnosis. PLoS One 2015;10:e0120911.

[26] Cullis PS, Gudlaugsdottir K, Andrews J. A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. PLoS One 2017;12:e0175213.
[27] Saleh AA, Ratajeski MA, Bertolet M. Grey Literature Searching for Health Sciences Systematic Reviews: A Prospective Study of Time Spent and Resources Utilized. Evid Based Libr Inf Pract 2014;9:28-50.
[28] Atakpo P, Vassar M. Publication bias in dermatology systematic reviews and meta-analyses. J Dermatol Sci 2016;82:6974.
[29] Anderson L, Taylor RS. Cardiac rehabilitation for people with heart disease: an overview of Cochrane systematic reviews. Cochrane database Syst Rev 2014;12:CD011273.
[30] Ho RST, Wu X, Yuan J, Liu S, Lai X, Wong SYS, et al. Methodological quality of meta-analyses on treatments for chronic obstructive pulmonary disease: a cross-sectional study using the AMSTAR (Assessing the Methodological Quality of Systematic Reviews) tool. NPJ Prim care Respir Med 2015;25:14102.
[31] Han JL, Gandhi S, Bockoven CG, Narayan V, Dahm P. The landscape of systematic reviews in urology (1998 through 2015): An assessment of methodologic quality. BJU Int 2017;119:638-49.
[32] Martel G, Duhaime S, Barkun JS, Boushey RP, Ramsay CR, Fergusson DA. The quality of research synthesis in surgery: the case of laparoscopic surgery for colorectal cancer. Syst Rev 2012;1:14.
[33] Hersi M, Quach P, Wang M-D, Gomes J, Gaskin J, Krewski D. Systematic reviews of factors associated with the onset and progression of neurological conditions in humans: A methodological overview. Neurotoxicology 2016 Jul 1. pii: S0161-813X(16)30118-8.
[34] Brown J, Farquhar C. Endometriosis: an overview of Cochrane Reviews. Cochrane database Syst Rev 2014;3:CD009590.
[35] Fernandes RM, Cary M, Duarte G, Jesus G, Alarcao J, Torre C, et al. Effectiveness of needle and syringe Programmes in people who inject drugs - An overview of systematic reviews. BMC Public Health 2017;17:309.
[36] Nagasawa DT, Bui TT, Lagman C, Lee SJ, Chung LK, Niu T, et al. Isolated Transverse Process Fractures: A Systematic Analysis. World Neurosurg 2017;100:33641.
[37] Viswanathan M, Treiman KA, Kish-Doto J, Middleton JC, Coker-Schwimmer EJL, Nicholson WK. Folic Acid Supplementation for the Prevention of Neural Tube Defects: An Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2017;317:190203.
[38] Hamill J, Rahiri JL, Hill AG. Analgesic effect of intraperitoneal local anesthetic in surgery: an overview of systematic reviews. J Surg Res. 2017;(212):16777.

[39] Papageorgiou PN, Deschner J, Papageorgiou SN. Effectiveness and Adverse Effects of Deep Brain Stimulation: Umbrella Review of Meta-Analyses. J Neurol Surg A Cent Eur Neurosurg 2017;78:18090.
[40] Barbateskovic M, Larsen LK, Oxenboll-Collet M, Jakobsen JC, Perner A, Wetterslev J. Pharmacological interventions for delirium in intensive care patients: a protocol for an overview of reviews. Syst Rev 2016;5:211.
[41] Dombrowski SU, Campbell P, Frost H, Pollock A, McLellan J, MacGillivray S, et al. Interventions for sustained healthcare professional behaviour change: a protocol for an overview of reviews. Syst Rev 2016;5:173.
[42] Morden A, Horwood J, Whiting P, Savovic J, Tomlinson L, Blakeman T, et al. The risks and benefits of patients temporarily discontinuing medications in the event of an intercurrent illness: a systematic review protocol. Syst Rev 2015;4:139.
[43] Mathur S, Conway DI, Worlledge-Andrew H, Macpherson LMD, Ross AJ. Assessment and prevention of behavioural and social risk factors associated with oral cancer: protocol for a systematic review of clinical guidelines and systematic reviews to inform Primary Care dental professionals. Syst Rev 2015;4:184.
[44] PROSPERO: International prospective register of systematic reviews [Internet]. [cited 2016 Jan 1]. Available from: http://www.crd.york.ac.uk/prospero/
[45] Jean-Luc F, Ferrer P, Gouverneur AD, Berkemeyer S, Vidal X, et al. A new risk of bias checklist applicable to randomized trials, observational studies and systematic reviews was developed and validated to be used for systematic reviews focusing on drug adverse events. J Clin Epidemiol 2017;pii: S0895-4356(16)30582-0.
[46] Burda BU, Holmer HK, Norris SL. Limitations of A Measurement Tool to Assess Systematic Reviews (AMSTAR) and suggestions for improvement. Syst Rev 2016;5:110.

Discusión

En resumen, desde una perspectiva general, este trabajo ha permitido comparar la magnitud, dirección y calidad de la evidencia existente sobre la eficacia y seguridad a corto plazo de los fármacos biológicos aprobados para el tratamiento de la psoriasis en placas moderada-severa en adultos, así como la calidad metodológica y el riesgo de sesgo de las revisiones sistemáticas realizadas sobre esta enfermedad durante los últimos treinta años.

Tomar decisiones es el acto central de la práctica médica. Son múltiples los resultados de estudios llevados a cabo con datos primarios y secundarios que sirven de base para la realización de recomendaciones. Y cada evidencia obtenida debería ayudar a la producción de nuevas de mayor calidad. Sin embargo, recientemente un manifiesto por la ciencia [130] cuestiona la reproducibilidad de las investigaciones publicadas poniendo en duda sus resultados.

Nuestro proyecto ha tratado de servirse de las mejores pruebas científicas para disminuir la incertidumbre existente en el campo de estudio de la psoriasis con el propósito de optimizar las decisiones que incumben la salud de estos pacientes. Para ello hemos puesto todo el empeño en desarrollar las estrategias de investigación más apropiadas, centradas en la síntesis de la evidencia, aplicando el mejor método científico y con la mayor transparencia de la que hemos sido capaces. El fin último, además de mejorar el conocimiento existente, es facilitar que nuestro trabajo pueda ser reproducido y que los esfuerzos realizados puedan servir para generar nuevas preguntas de investigación o contribuyan a responder las ya existentes.

Una vez expuesto el compendio de artículos científicos a los que ha dado lugar el presente proyecto de investigación, nos gustaría poner en valor el hecho de que nuestros resultados pueden ser reproducidos por cualquier investigador. Para ello, hemos publicado en PROSPERO un protocolo de cada de una de las investigaciones con la metodología seguida y las herramientas empleadas; hemos aportado el listado de las revisiones incluídas y excluídas, los datos extraídos de los documentos y los *scripts* con el código de Stata y de R utilizados para su análisis y visualización; además, hemos añadido, como documentos adicionales a esta tesis, las respuestas a las cuestiones planteadas por editores y revisores de

las revistas a las que se ha enviado los manuscritos originales para su posible publicación. Estimamos que las aportaciones de estos investigadores anónimos han sido clave para mejorar la investigación realizada, por lo que hemos creído apropiado ofrecerlas al lector para su consideración. En definitiva, consideramos que todos estos elementos aportan pruebas adicionales de gran valor, en cierto modo novedosas en este tipo de proyectos, para que el tribunal pueda tenerlo en cuenta al evaluar el grado de capacitación en investigación alcanzado por el doctorando.

Además, desde su reciente publicación, los artículos han suscitado gran interés en la comunidad científica, dando lugar a una carta al director, un editorial, varias citas en artículos científicos y una rápida repercusión en medios de difusión en internet, medida a través del índice Altmetrics. Destacamos también la reciente incorporación de los investigadores que han contribuído en este proyecto como miembros de dos grupos de trabajo de Cochrane: tanto en el grupo dedicado a las enfermedades de la piel (Cochrane Skin Group, CSG) como en el de metodología y riesgo de sesgo (Cochrane Bias Methods Group, BMG), que abordan aspectos de investigación que han sido nucleares en este trabajo.

Además, queremos resaltar la imprescindible labor del equipo de investigadores que han participado en la realización del presente proyecto. El valor de su aportación no solamente se justifica por ser necesario para cumplir con los estándares más altos de investigación disponer de más de un revisor para la selección de los estudios, la extracción de los datos, la evaluación del riesgo de sesgo o el análisis de la calidad metodológica, o porque sea deseable el conocimiento específico de otros investigadores en algunos de los temas o procedimientos, sino que su colaboración como grupo desvela que la socialización de la investigación la enriquece en contenido y en calidad. Un trabajo de este tipo, dividido en conjuntos de procesos, lleva implícita la emisión de múltiples juicios individuales cuya síntesis debe alcanzarse por acuerdo. En cierta medida, trabajar en el seno de un grupo de investigación permite "tomar decisiones sin ser mecanicistas", ya que acordar requiere modificar, ajustar o acomodar puntos de vista personales a los del conjunto, que es la intención, el espíritu que se promulga en las guías publicadas para la realización de este tipo de estudios.

La repercusión de cualquier hallazgo de una investigación debe ser considerada en función de los medios empleados en alcanzarlo. De esta forma, aunque los resultados obtenidos son interesantes porque abordan las pruebas existentes sobre una enfermedad prevalente y asociada a elevada morbimortalidad, que produce un gran impacto en la calidad de vida de los pacientes, consideramos que su especial relevancia radica en la estrategia empleada para su desarrollo, siempre guiada por el cumplimiento de los estándares más altos de calidad metodológica.

La aplicación de los últimos conceptos surgidos de la experiencia empírica aplicados a la síntesis de la evidencia deben acortar cada vez más la distancia entre el resumen del conocimiento y la realidad. En este sentido, las estimaciones de los efectos deben considerarse como una aproximación al efecto real y sin condicionar su importancia a la magnitud del efecto en sí, ni a la significación estadística asociada. La pirámide de la evidencia es una representación de metodologías que, desde su base hasta el punto mas alto, las revisiones sistemáticas, tienen el potencial de producir un conocimiento cada vez más cercano a la realidad desconocida. Sin embargo es el rigor metodológico, la capacidad de evitar errores sistemáticos y la adaptación a los nuevos conceptos demostrados por la evidencia en la realización de cada tipo de investigación, lo que en su conjunto debe ser tenido en cuenta cuando se consideran los resultados para establecer las conclusiones.

Sin embargo, la credibilidad en los hallazgos se pone en muchas ocasiones en duda en los últimos años, en los que la producción de revisiones sistemáticas se ha disparado, siendo muchas de ellas innecesarias, engañosas y conflictivas, cuyos fines, más que mejorar el conocimiento, parecen responder a intereses alejados de los de la ciencia . La dificultad para identificar en este océano de datos las mejores pruebas de revisiones de alta calidad metodológica y bajo riesgo de sesgo es una barrera que puede limitar la adopción de esta nueva ciencia, en favor del uso de criterios que asocian mayor incertidumbre y sesgo, como los basados en la experiencia no sistemática, la opinión de expertos o el razonamiento fisiopatológico. Es decir, el contexto de producción masiva de documentos de síntesis de baja utilidad pone en riesgo el nuevo paradigma de basar la medicina en la evidencia para aumentar la certidumbre en el uso de los medios disponibles. La comunidad científica debe, por tanto, esforzarse en desarrollar y seguir mejorando las herramientas que contribuyen a interpretar el verdadero alcance de las investigaciones, aumentando su difusión y replicando con ellas resultados obtenidos previamente mediante otros procedimientos. La segunda parte de nuestro proyecto ha estado centrada en la investigación meta-epidemiológica, es decir, en la evaluación de la calidad de este tipo de evidencia publicada sobre la psoriasis y la obtención de modelos predictivos de la misma. Sin duda, poder realizar un juicio sustentado en la evidencia sobre el grado de certidumbre que contiene una investigación es uno de los hitos que mejor defienden la aplicación de los hallazgos de esta nueva ciencia.

Cada uno de los trabajos realizados que aportamos en este compendio ha sido diseñado *a priori*. Además, el carácter prospectivo de los mismos ha sido refrendado por el cumplimiento con las normas establecidas por PROSPERO. Este proceder aporta transparencia y elimina el riesgo de sesgo que las decisiones *ad hoc* suponen para cualquier investigación. Encontrar la forma de hacer extensible esta manera de proceder para cualquier estudio de este tipo debe convertirse en un objetivo prioritario de las organizaciones involucradas en

la mejora de la calidad y la transparencia de la investigación, como Equator network, y de editores y revisores de las revistas científicas, que deberían ayudar a exigir y fomentar su cumplimiento.

Sobre la base anterior, este proyecto está dedicado a la investigación del estándar de la síntesis de la evidencia sobre psoriasis empleando para ello las aproximaciones epistemológicas descritas para su estudio desde dos puntos de vista: la realización de revisiones sistemáticas y meta-análisis en red, que comprende tanto la conducción como la notificación; y la evaluación meta-epidemiológica de la evidencia recogida en las revisiones sistemáticas y meta-análisis, empleando para ello la metodología de conducción y notificación descritas.

Con respecto a la primera parte del proyecto, los resultados de nuestro trabajo permitirían el desarrollo de guías de práctica clínica que ayuden a médicos, pacientes y gestores en salud a adoptar las mejores medidas para el tratamiento de la psoriasis. Esto es así no solamente porque incluye todas las alternativas terapeúticas autorizadas en el momento de su realización, sino especialmente porque permite analizar las diferencias entre ellas comparando el equilibrio entre el riesgo y el beneficio estimados y aportando una estimación del grado de certidumbre contenido en cada uno de los hallazgos. Todo ello se ha llevado a cabo siguiendo las últimas guías y procedimientos metodológicos para su conducción y notificación, y empleando recursos de financiación pública competitiva que hemos obtenido a través del Instituto de Salud Carlos III para su realización.

La puesta al día de este tipo de conocimiento es, a su vez, reto y necesidad, tanto de los productores como de los consumidores de revisiones sistemáticas. En ambos casos, productores y consumidores se enfrentan a un ingente número de trabajos de investigación que continuamente se publican y que desbordan la capacidad de clínicos y evaluadores para elaborar un juicio tras su lectura crítica. Sin embargo, dicha limitación en tiempo y recursos probablemente pueda ser superada en un futuro con la incoporación de tecnologías que implementan en el flujo de trabajo procesos de automatización para la realización o la evaluación de una revisión sistemática. Apoyar la aplicación de nuevas estrategias basadas en recursos innovadores de análisis computacional (*text mining*, machine learning y *deep learning*) a la síntesis de la evidencia parece el mejor camino en la actualidad para solventar estas limitaciones.

El estudio de cada una de las herramientas meta-epidemiológicas empleadas ofrece una perspectiva de la dificultad de su desarrollo debido a la ausencia de un *gold standar* con el que compararse. Sus dificultades radican en que deben medir un aspecto del que no se conoce su magnitud ni dirección, empleando una metodología que no ha sido bien desarrollada hasta el momento. Sin embargo sus fundamentos se basan en pruebas empíricas y cuentan con la experiencia del desarrollo de los conceptos de la síntesis de la evidencia. Como

ocurre en ciencia, cada refutación del conocimiento acerca cada vez más la estimación a la realidad. Y cada prueba producida debe servir para la construcción de este camino. La herramienta de la Colaboración Cochrane para el riesgo de sesgo, la metodología de GRADE para la gradación de la calidad de la evidencia, las herramientas AMSTAR y ROBIS para la evaluación de la calidad metodológica y el análisis del riesgo de sesgo de las revisiones sistemáticas, respectivamente, son ejemplos claros de los instrumentos meta-epidemiológicos empleados en nuestros proyecto que contribuyen a dicho fin.

El uso de la herramienta de la Cochrane aplicada a los ensayos clínicos aleatorizados de los fármacos evaluados nos ha enseñado que, a pesar del prestigio de este tipo de diseño, existen limitaciones inherentes a su uso en psoriasis, como es el caso de la subjetividad de las medidas evaluadas, que ponen en riesgo la validez de la respuestas a sus preguntas de investigación. La ausencia de diseños centrados en seguridad o el manejo de la síntesis de los datos son aspectos que sin duda necesitan ser abordados y mejorados en el futuro para evitar desviaciones sistemáticas del efecto real.

Nuestro meta-análisis en red ha permitido evaluar por primera vez la calidad de los resultados de las medidas de efecto y seguridad a corto plazo de los fármacos biológicos empleando la metodología GRADE. Incorporando el grado de calidad de la evidencia a la magnitud y dirección de la estimación de los efectos, permite tener una perspectiva mas clara de las espectativas de uso de estos tratamientos. Los resultados obtenidos sobre la calidad de la evidencia arrojan incertidumbres que son mayores sobre los aspectos de seguridad que sobre los de eficacia. Esta diferencia estimula a reflexionar qué motivo debe existir para que los estudios primarios aporten una mejor calidad de los datos cuando dos principios activos se enfrentan entre sí que cuando uno lo hace con el placebo. Si existe una causa que produzca una desviación de la calidad sistemática o si ésta es debida únicamente al azar, en estos casos, es una pregunta cuya respuesta puede enriquecer la evaluación del alcance de estas investigaciones. Por el contrario, el hecho de que los resultados recientes mejoren la confianza de los más antiguos impulsa el recorrido de la medicina basada en la evidencia.

La necesidad de mantener una actitud de excepticismo *a priori* sobre los resultados y conclusiones de los documentos de síntesis de la evidencia, independientemente de la fuente que los publica, es el concepto más importante surgido del manejo de la herramienta AMSTAR en nuestro proyecto. A pesar de que puedan ponerse en cuestión aspectos sobre la construcción y aplicación de dicho instrumento de evaluación, su uso arroja claridad sobre lo realmente trascendente en ciencia: poner el foco en el método mas allá de la magnitud y dirección de los resultados.

Buscar más allá de los datos (metadatos) del informe científico de los manuscritos nos ha dado la oportunidad de poder profundizar mejor en los factores que dirigen nuestros resultados y las relaciones no explícitas entre ellos, lo que nos ha ayudado a predecirlos mediante modelos matemáticos. Sin embargo, creemos necesario que puedan ser replicados llevando a cabo evaluaciones similares a las nuestras en otras áreas temáticas. Las principales razones que condicionan las diferencias encontrada en la calidad metodológica obedecen a tres clases de factores: los intereses no científicos de los autores, las limitaciones existentes para la comunicación correcta de la investigación realizada y los relacionados con las revistas científicas encargadas de su difusión. El tipo de financiación de los trabajos analizados o la presencia de autores con conflictos de intereses como factores condicionantes de la calidad metodológica ya habían sido descritos previamente en la literatura. Sobre la limitación existente para poder plasmar en un espacio limitado, de forma explícita y transparente los aspectos más importantes del desarrollo de la investigación, hemos de tener presente que se han desarrollado protocolos para la notificación de los trabajos cuyo cumplimiento aumenta la utilidad de los mismos. Finalmente, a pesar de que la relevancia y el modo en que se mide el impacto que ejerce una publicación en la comunidad científica siguen siendo temas que suscitan discusión, nosotros hemos encontrado que existe relación entre la repercusión de los resultados, las revistas y la calidad científica de sus trabajos.

Realizar una evaluación meta-epidemiológica es un proceso complejo, que requiere del conocimiento y manejo actualizado de las herramientas desarrolladas. Esto puede suponer un impedimento para su uso generalizado. Sin embargo sería deseable que los difusores del conocimiento tuvieran en cuenta la incorporación de estas herramientas cuando lleven a cabo los procesos de selección de la literatura publicada. De modo que a los consumidores de la misma les fuera posible reconocer el grado de confianza en los hallazgos y conclusiones.

Una de las conclusiones más interesantes de nuestro trabajo ha sido que el hecho de evaluar las desviaciones sistemáticas de los documentos de síntesis va más allá de la evaluación de la calidad metodológica de los mismos. En este sentido, el instrumento ROBIS recorre los aspectos del rigor de la conducción de las revisiones sistemáticas y cómo han sido consideradas sus limitaciones, las de la evidencia aportada por el conjunto de estudios que la integran y por la tendencia a comunicar resultados basándose en la significación estadística de los mismos. Hasta el momento, existe poca experiencia del uso de esta herramienta y, aunque su construcción y preguntas que la integran están dirigidas a examinar el riesgo de sesgo, es necesario que la evidencia empírica corrobore que realmente es ésto lo que evalúa.

Finalmente, este trabajo ha aportado evidencia a la discusión entre calidad metodológica y riesgo de sesgo . En este sentido, aún pueden encontrarse entre los ítems que apoyan la redacción y el cumplimiento prospectivo los de protocolos de las revisiones sistemáticas, cómo el de PROSPERO de la Universidad de York, la consideración de la relación de igualdad de ambos conceptos, en los que calidad metodológica aparece entre paréntesis adyacente al riesgo de sesgo, cuando se cuestiona sobre la evaluació en de la validez interna de los estudios incluidos en este tipo de investigaciones. Probablemente este tipo de hallazgos, en los que una gran proporción de revisiones sistemáticas de alta calidad muestran alto riesgo de sesgo, puedan ayudar a eliminar esta confusión y aclarar estos conceptos.

Conclusiones

(i) **Infliximab 5 mg/Kg cada 8 semanas y secukinumab 300 mg cada 4 semanas fueron los agentes más eficaces en comparación con placebo en relación con los resultados de PASI 75 y PASI 90, respectivamente, obtenidos en la semana 10-16.** Desprendida de:Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

(ii) **Infliximab 5 mg/Kg cada 8 semanas y secukinumab 300 mg cada 4 semanas fueron los agentes con más riesgo de presentar al menos un efecto adverso y al menos una infección, respectivamente.** Desprendida de:Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

(iii) **Ustekinumab 90 mg cada 12 semanas resultó ser el tratamiento con mejor perfil riesgo beneficio para las medidas estudiadas, ocupando el tercer lugar en el ranking de eficacia y en posición similar al placebo en relación al ranking de acontecimientos adversos.** Desprendida de: Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

(iv) **Los ensayos clínicos de fármacos biológicos autorizados en el tratamiento en psoriasis presentan globalmente bajo riesgo de sesgo si bien existen incertidumbres para la mayoría de fuentes de sesgo. Aunque no es posible conocer con certeza**

la dirección del sesgo cuyo riesgo es introducido por dichas incertidumbres, la evidencia empírica demuestra que estas fuentes de sesgo aumentan el riesgo de sobreestimación de los efectos de la intervención, lo que es necesario tener en cuenta para la correcta valoración de la magnitud de los efectos de la intervención. Desprendida de: Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

(v) **En relación a la calidad de la evidencia:**

- Es globalmente mejor para los resultados de eficacia que para los de seguridad.
  - En relación a los resultados de eficacia, varía entre los diferentes fármacos y dosis, siendo alta para ustekinumab, moderada para infliximab, secukinumab y etanercept, y baja para adalimumab.
  - En relación a los resultados de seguridad, en general la calidad de la evidencia es baja o muy baja. Es probable que nuevos estudios modifiquen la confianza en la estimación del efecto y su magnitud para la mayoría de los tipos de acontecimientos adversos medidos y de los ensayos clínicos, existiendo en este sentido incertidumbre sobre la seguridad a corto plazo de estos fármacos.
- Es mejor para los resultados de ensayos *head to head* que aquellos que comparan un fármaco con placebo.
- Es mejor para los ensayos clínicos más recientes que para los más antiguos.

Desprendida de: Gómez-García F*, Epstein D*, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol.* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.

(vi) **La calidad metodológica de las revisiones sistemáticas y los meta-análisis publicados sobre psoriasis es subóptima. Únicamente un 17% de las revisiones sistemáticas evaluadas es de alta calidad empleando AMSTAR. Los aspectos con menor porcentaje de complimiento fueron: la presentación de una lista de estudios excluidos con el motivo de exclusión, la evaluación del sesgo de publicación, la inclusión de literatura gris y la definición de un diseño *a priori*.**

(vii) **Nuestro modelo predice una alta calidad metodológica cuando una revisión sistemática incluye uno o varios de los siguientes factores:**

- Un meta-análisis.
- Ha sido financiado por instituciones académicas
- El número de autores con conflicto de intereses es bajo.
- El factor de impacto de la revista donde ha sido publicado es alto.
- El número de páginas de artículos es elevado.

Desprendida de: Gómez-García F*, Ruano J*, Aguilar-Luque M, Gay-Mimbrera J, Maestre-Lopez B, Sanz-Cabanillas JL, Carmona-Fernández PJ, González-Padilla M, Vélez García-Nieto A, Isla-Tejera B. **Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest and bibliometric indices as predictors of methodological quality.** *Br J Dermatol.* 2017 Jun;176(6):1633-1644. doi: 10.1111/bjd.15380. Epub 2017 May 19.

(viii) **El análisis de componente principal y clustering ha permitido encontrar más subgrupos significativos de revisiones sistemáticas que los establecidos simplemente al emplear los niveles bajo, medio y alto derivados de la puntuación global obtenida con AMSTAR.** Desprendida de: Gómez-García F*, Ruano J*, Aguilar-Luque M, Gay-Mimbrera J, Maestre-Lopez B, Sanz-Cabanillas JL, Carmona-Fernández PJ, González-Padilla M, Vélez García-Nieto A, Isla-Tejera B. **Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest and bibliometric indices as predictors of methodological quality.** *Br J Dermatol.* 2017 Jun;176(6):1633-1644. doi: 10.1111/bjd.15380. Epub 2017 May 19.

(ix) **El riesgo de sesgo de las revisiones sistemáticas y los meta-análisis publicados sobre psoriasis fue alto. Únicamente el 14% de las revisiones sistemáticas presentó un bajo riesgo de sesgo según ROBIS. Las preguntas de señalización que la mayoría de los estudios no recogieron tenían relación con si los autores habían obtenido los datos relevantes para su uso en la síntesis o si de algún modo aclaraban la robustez de los resultados en función de diagramas de embudo y análisis de sensibilidad.** Desprendida de: Gómez-García F*, Ruano J*, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, Maestre-López B, Carmona-Fernández PJ, González-Padilla M, García-Nieto AV, Isla-Tejera B. **Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool.** *J Clin Epidemiol.* 2017 Sep 9. pii: S0895-4356(17)30534-6. doi: 10.1016/j.jclinepi.2017.08.015. [Epub ahead of print]

(x) **En nuestro estudio, la mayoría de las revisiones sistemáticas y meta-análisis sobre intervenciones en psoriasis que fueron clasificadas como de alta y moderada calidad metodológica según AMSTAR, mostraron un alto riesgo de sesgo con ROBIS. Dado que ambas herramientas han demostado validez interna, estos hallazgos aportan una base científica a la discusión sobre los conceptos calidad metodológica y riesgo de sesgo. Por ello, se podrían considerar complementarias en la evaluación de la calidad científica de las revisiones sistemáticas.** Desprendida de: Gómez-García F*, Ruano J*, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, Maestre-López B, Carmona-Fernández PJ, González-Padilla M, García-Nieto AV, Isla-Tejera B. **Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool.** *J Clin Epidemiol.* 2017 Sep 9. pii: S0895-4356(17)30534-6. doi: 10.1016/j.jclinepi.2017.08.015. [Epub ahead of print]

Bibliografía

[1] Beyer V,Wolverton SE, *Recent trends in systemic psoriasis treatment costs.* Archives of Dermatology 2010,1 (146): 46-54

[2] Pathirana D , Ormerod , and Saiag P , et al, *European S3-guidelines on the systemic treatment of psoriasis vulgaris.* Eur Acad Dermatol Venereol. 2010 Dec;24(12):1458-67. doi: 10.1111/j.1468-3083.2010.03671.x.

[3] Ogdie A , Yu Y , Haynes K , Love TJ , Maliha S ,Jiang Y , Troxel AB ,Hennessy S , Kimmel SE ,Margolis DJ , Choi H ,Mehta NN, Gelfand JM, *Risk of major cardiovascular events in patients with psoriatic arthritis, psoriasis and rheumatoid arthritis: a population-based cohort study.* Ann Rheum Dis. 2015 Feb;74(2):326-32. doi: 10.1136/annrheumdis-2014-205675. Epub 2014 Oct 28.

[4] Krueger G , Koo J , Lebwohl M , Menter A, Stern RS , Rolstad T., *The impact of psoriasis on quality of life: results of a 1998 National Psoriasis Foundation patient-membership survey.* Arch Dermatol. 2001 Mar;137(3):280-4.

[5] Chi CC , Wang SH., *Efficacy and cost-efficacy of biologic therapies for moderate to severe psoriasis: a meta-analysis and cost-efficacy analysis using the intention-to-treat principle.* Biomed Res Int. 2014;2014:862851. doi: 10.1155/2014/862851. Epub 2014 Jan 29.

[6] Cochrane A, *Effectiveness and efficiency: Random reflections on health services* . Nuffield Trust.

[7] Abuabara K , Freeman EE , Dellavalle R., *The role of systematic reviews and meta-analysis in dermatology.* J Invest Dermatol. 2012 Nov;132(11):e2. doi: 10.1038/jid.2012.392.

[8] Higgins JPT, Green S (editors)., *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated Sept 201].* The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

[9] Atkins D and Fink K and Slutsky J., *Better information for better health care: the Evidence-based Practice Center program and the Agency for Healthcare Research and Quality.].* Ann Intern Med. 2005 Jun 21;142(12 Pt 2):1035-41. Review.

[10] Akers J , Aguiar-Ibañez R, Sari A , Beynon Sand, Booth A , Burch J , Chambers D , Craig D , Dalton J , Duffy S and Eastwood A , Fayter D, Fonseca T , Fox D, Glanville J , Golder S , Hempel S and GW K., *Centre for Reviews and Dissemination (2009) Systematic reviews: CRD's guidance for undertaking reviews in health care [Internet]. [cited 2017 Sept 6].* York University, York,2009.

[11] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF., *Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses.* Lancet. 1999 Nov 27;354(9193):1896-900. Review.

[12] Stroup DF , Berlin JA , Morton SC , Olkin I , Williamson GD , Rennie D , Moher D , Becker BJ , Sipe TA , Thacker SB., *Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group.* JAMA. 2000 Apr 19;283(15):2008-12. Review.

[13] Welch V , Petticrew M , Tugwell P , Moher D , O'Neill J , Waters E , White H and PRISMA-Equity Bellagio group., *PRISMA-Equity 2012 extension: reporting guidelines for systematic reviews with a focus on health equity.* PLoS Med. 2012;9(10):e1001333. doi: 10.1371/journal.pmed.1001333. Epub 2012 Oct 30.

[14] Ioannidis JP., *The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses.* Milbank Q. 2016 Sep;94(3):485-514. doi: 10.1111/1468-0009.12210.

[15] Reddi A, Prescott L , Doney E , Delamere F, Kollipara R, Dellavalle RP and Williams HC., *The Cochrane Skin Group: a vanguard for developing and promoting evidence-based dermatology.* J Evid Based Med. 2013 Nov;6(4):236-42. doi: 10.1111/jebm.12068.

[16] Collier A , Heilig L , Schilling L , Williams H , Dellavalle RP., *Cochrane Skin Group systematic reviews are more methodologically rigorous than other systematic reviews in dermatology.* Br J Dermatol. 2006 Dec;155(6):1230-5.

[17] Davila-Seijo P and Batalla A and Garcia-Doval I., *Ussefulness of Cochrane Skin Group reviews for clinical practice.* Actas Dermosifiliogr. 2013 Oct;104(8):679-84. doi: 10.1016/j.adengl.2012.12.009. Epub 2013 Aug 13.

[18] Owen CM and Chalmers RJ and O'Sullivan T and Griffiths CE., *Antistreptococcal interventions for guttate and chronic plaque psoriasis.* Cochrane Database Syst Rev. 2000;(2):CD001976. Review.

[19] Williams H., *Why is the center of evidence-based dermatology relevant to Indian dermatology?.* Indian J Dermatol. 2009;54(2):118-23. doi: 10.4103/0019-5154.53184.

[20] Ching Chi-Chi., *Evidence-based dermatology* . Dermatologica Sinica. 2013;31(1):2-6.

[21] Bell HK and Ormerod AD and BAD Therapy and Guidelines Subcommittee., *Writing a British Association of Dermatologists clinical guideline: an update on the process and guidance for authors.* Br J Dermatol. 2009 Apr;160(4):725-8. doi: 10.1111/j.1365-2133.2008.09026.x.

[22] Guyatt G.H. , Haynes R.B. , Jaeschke R.Z. , Cook D.J. , Green L ,Naylor C.D ,and Wilson M.C. , Richardson W.S. , *Users' Guides to the Medical Literature: XXV. Evidence-Based Medicine: Principles for Applying the Users' Guides to Patient Care. Evidence-Based Medicine Working Group.* JAMA. 2000 Sep 13;284(10):1290-6.

[23] Sackett DL ,Rosenberg WM , Gray JA ,Haynes RB , Richardson WS., *Evidence based medicine: what it is and what it isn't.* BMJ. 1996 Jan 13;312(7023):71-2.

[24] Sackett DL, Straus SE, Richardson WS , Rosemberg W , Haynes B., *Evidence based medicine: what it is and what it isn't.* Churchill Livingstone.Edinburgh.2000

[25] Kohatsu ND , Robinson JG , Torner JC. , *Evidence-based public health: an evolving concept.* Am J Prev Med. 2004 Dec;27(5):417-21. Review.

[26] Hoffman T, Bennett S, Del Mar, C., *Evidence-based public health: an evolving concept.* Evidence-Based Practice: Across the Health Professions(2nd ed.).2013. Chatswood, NSW: Elsevier.

[27] DiCenso A, Bayley L, Haynes RB., *Editorial: Accessing preappraised evidence: fine-tuning the 5S model into a 6S model.* Ann Intern Med. 2009 Sep 15;151(6):JC3-2, JC3-3

[28] Oxman AD, Guyatt GH.., *The science of reviewing research.* Ann N Y Acad Sci. 1993 Dec 31;703:125-33; discussion 133-4. Review.

[29] Moher D , Liberati A , Tetzlaff J , Altman DG and The PRISMA Group (2009)., *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The*

PRISMA Statement. Ann Intern Med. 2009 Aug 18;151(4):264-9, W64. Epub 2009 Jul 20.

[30] Chandler J. , Churchill R. ,Higgins J., Lasserson T, Tovey D. The Cochrane Collaboration., *Methodological Expectations of Cochrane Intervention Reviews (MECIR): methodological standard for the conduct of new Cochrane Intervention Reviews.* Cochrane Editorial Unit, editor. Methodological Expectations of Cochrane Intervention Reviews (MECIR) 2013 Retrieved from http://www.editorial-unit.cochrane.org/sites/editorial

[31] Light RJ and Pillemer DB., *Summing Up: The Science of Reviewing Research..* Harvard University Press.1984

[32] Davies K., *Formulating the Evidence Based Practice Question: A Review of the Frameworks..* Evidence Based Library and Information Practice.2011.6 (2).

[33] Methley AM , Campbell S , Chew-Graham C, McNally R , Cheraghi-Sohi S., *PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews.* BMC Health Services Research201414:579.

[34] Egger M, Smith GD , Altman DG editors., *Systematic reviews in healthcare: Meta-analysis in context. 2nd edition.* BMJ Publishing Group.2001.London.

[35] Deeks JJ., *Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes.* Stat Med. 2002 Jun 15;21(11):1575-600.

[36] Khan KS , ter Riet G , Clanville J , Sowden AJ , Kleijnen J editors., *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews (CRD Report Number 4) (2nd edition).* NHS Centre for Reviews and Dissemination, University of York.2001.

[37] Chang SM , Matchar DB, Smetana GW , Umscheid CA editors., *Methods Guide for Medical Test Reviews.* Agency for Healthcare Research and Quality (US),Rockville (MD).2012.

[38] Deville WL and Bezemer PD and Bouter LM., *Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy..* J Clin Epidemiol. 2000 Jan;53(1):65-9.

[39] Wilczynski NL and Haynes RB and Hedges Team., *EMBASE search strategies for identifying methodologically sound diagnostic studies for use by clinicians and researchers.* BMC Med. 2005 Mar 29;3:7.

[40] Whiting P, Westwood M , Burke M , Sterne J , Glanville J.., *Systematic reviews of test accuracy should search a range of databases to identify primary studies.* J Clin Epidemiol. 2008 Apr;61(4):357-364. doi: 10.1016/j.jclinepi.2007.05.013.

[41] Edwards P , Clarke M , DiGuiseppi C , Pratap S , Roberts I, Wentz R., *Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records.* Stat Med. 2002 Jun 15;21(11):1635-40.

[42] Cooper HM and Ribble RG ., *Influences on the outcome of literature searches for integrative research reviews.* Knowledge.1989.10:179-201

[43] Jadad AR,Moore RA ,Carroll D ,Jenkinson C ,Reynolds DJ ,Gavaghan DJ, McQuay HJ.., *Assessing the quality of reports of randomized clinical trials: is blinding necessary?.* Control Clin Trials. 1996 Feb;17(1):1-12.

[44] Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA; Cochrane Bias Methods Group; Cochrane Statistical Methods Group., *The Cochrane Collaboration's tool for assessing risk of bias in randomised trials.* BMJ. 2011 Oct 18;343:d5928. doi: 10.1136/bmj.d5928.

[45] Simmonds MC, Higgins JP., *A general framework for the use of logistic regression models in meta-analysis.* Stat Methods Med Res. 2016 Dec;25(6):2858-2877. Epub 2014 May 12.

[46] Mavridis D ,Salanti G., *A practical introduction to multivariate meta-analysis.* Stat Methods Med Res. 2013 Apr;22(2):133-58. doi: 10.1177/0962280211432219.

[47] Lu G ,Ades AE., *A practical introduction to multivariate meta-analysis.* Stat Med. 2004 Oct 30;23(20):3105-24.

[48] Thompson SG ,Higgins JP., *How should meta-regression analyses be undertaken and interpreted?.* Stat Med. 2002 Jun 15;21(11):1559-73.

[49] Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group., *Systematic reviews of diagnostic test accuracy..* Ann Intern Med. 2008 Dec 16;149(12):889-97.

[50] Higgins JPT, Thompson SG, Deeks JJ,Altman DG.., *Measuring inconsistency in meta-analyses.* BMJ. 2003 Sep 6;327(7414):557-60. Review.

[51] Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR., *Publication bias in clinical research.* Lancet. 1991 Apr 13;337(8746):867-72.

[52] Dickersin K , Min YI , Meinert CL., *Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards.* JAMA. 1992 Jan 15;267(3):374-8.

[53] Lexchin J, Bero LA, Djulbegovic B, Clark O., *Pharmaceutical industry sponsorship and research outcome and quality: systematic review..* BMJ. 2003 May 31;326(7400):1167-70. Review.

[54] Hopewell S, McDonald S, Clarke M, Egger M., *Grey literature in meta-analyses of randomized trials of health care interventions.* Cochrane Database Syst Rev. 2007 Apr 18;(2):MR000010. Review.

[55] Abbasi K, *Compulsory registration of clinical trials.* BMJ. 2004 Sep 18;329(7467):637-8.

[56] Sterne JAC, Egger M., *Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis.* J Clin Epidemiol. 2001 Oct;54(10):1046-55.

[57] Terrin N, Schmid CH, Lau J., *In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias.* J Clin Epidemiol. 2005 Sep;58(9):894-901.

[58] Egger M, Davey Smith G, Schneider M, Minder C., *In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias.* BMJ. 1997 Sep 13;315(7109):629-34.

[59] Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ; GRADE Working Group. , *What is "quality of evidence" and why is it important to clinicians?.* BMJ. 2008 May 3;336(7651):995-8. doi: 10.1136/bmj.39490.551019.BE. Review.

[60] Egger M, Davey Smith G, Schneider M, Minder C., *In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias.* BMJ. 1997 Sep 13;315(7109):629-34.

[61] Parmelli E, Amato L, Saitto C, Davoli M; Gruppo di Lavoro "DECIDE Italia., *DECIDE: developing and evaluating communication strategies to support informed*

decisions and practice based on evidence. Recenti Prog Med. 2013 Oct;104(10):522-31. doi: 10.1701/1349.14997. Review.

[62] Parmelli E, Amato L, Saitto C, Davoli M; Gruppo di Lavoro "DECIDE Italia., *World Health Organization recommendations are often strong based on low confidence in effect estimates.*. J Clin Epidemiol. 2014 Jun;67(6):629-34. doi: 10.1016/j.jclinepi.2013.09.020. Epub 2014 Jan 3. Review.

[63] Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG., *Epidemiology and reporting characteristics of systematic reviews.*. PLoS Med. 2007 Mar 27;4(3):e78.

[64] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF., *Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses.* Lancet. 1999 Nov 27;354(9193):1896-900. Review.

[65] Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF., *Obtaining information accurately and quickly: Are structured abstract more efficient?.* J Infor Sci. 1996. 22:349-356

[66] Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, Gøtzsche PC, Lasserson T, Tovey D; PRISMA for Abstracts Group., *PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts.* PLoS Med. 2013;10(4):e1001419. doi: 10.1371/journal.pmed.1001419. Epub 2013 Apr 9.

[67] Counsell C., *Formulating questions and locating primary studies for inclusion in systematic reviews.* Ann Intern Med. 1997 Sep 1;127(5):380-7. Review.

[68] Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG., *Empirical evidence for selective reporting of outcomes in randomized trials: Comparison of protocols to published articles.* JAMA. 2004 May 26;291(20):2457-65.

[69] Zhang L, Sampson M, McGowan J, *Reporting of the role of expert searcher in Cochrane reviews.* Evid Based Libr Info Pract.2006. 1:3-16

[70] Jones AP, Remmington T , Williamson PR , Ashby D , Smyth RL., *High prevalence but low impact of data extraction and reporting errors were found in Cochrane systematic reviews.*. J Clin Epidemiol.2005.58:741-742

[71] Glasziou P ,Meats E ,Heneghan C , Shepperd S ., *What is missing from descriptions of treatment in trials and reviews?.* BMJ. 2008 Jun 28;336(7659):1472-4. doi: 10.1136/bmj.39590.732037.47.

[72] Song F,and Eastwood AJ, Gilbody S, Duley L, Sutton AJ., *Publication and related biases..* Health Technol Assess. 2000;4(10):1-115. Review.

[73] Lewis S, Clarke M., *Forest plots: Trying to see the wood and the trees..* BMJ. 2001 Jun 16;322(7300):1479-80.

[74] Jorgensen AW, Hilden J,Gotzsche PC ., *Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: Systematic review.* BMJ. 2006 Oct 14;333(7572):782. Epub 2006 Oct 6. Review.

[75] Jorgensen AW, Hilden J,Gotzsche PC ., *The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and Explanations.* Ann Intern Med. 2015 Jun 2;162(11):777-84. doi: 10.7326/M14-2385.

[76] Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. ., *More informative abstracts revisited..* Ann Intern Med. 1990 Jul 1;113(1):69-76. Review.

[77] Haynes RB, Mulrow CD, Huth EJ, Altman DG, Gardner MJ. ., *Reporting of trials presented in conference abstracts needs to be improved..* J Clin Epidemiol. 2006 Jul;59(7):681-4.

[78] Brownson RC , Baker EA , Leet TL , Gillespie KN , True WR ., *Evidence-Based Public Health. 2nd ed.* Oxford University Press.New York.2011.

[79] Roever-Borges LS. , *Understanding Meta-Epidemiological Studies.* International Journal of Cardiovascular Sciences. 2016;29(4):326-328

[80] Pildal J, Chan AW, Hróbjartsson A , Forfang E , Altman , Gøtzsche PC. , *Understanding Meta-Epidemiological Studies.* BMJ. 2005 May 7;330(7499):1049. Epub 2005 Apr 7. Review.

[81] Zhang W. , *Meta-epidemiology: building the bridge from research evidence to clinical practice.* Osteoarthritis Cartilage.2010.18;Suppl 2:S1

[82] Chaimani A, Vasiliadis HS, Pandis N, Schmid CH, Welton NJ, Salanti G. , *Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study.* Int J Epidemiol. 2013 Aug;42(4):1120-31. doi: 10.1093/ije/dyt074. Epub 2013 Jun 27. Review.

[83] Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. , *Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists* Control Clin Trials. 1995 Feb;16(1):62-73.

[84] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. , *Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews.* BMC Med Res Methodol. 2007 Feb 15;7:10.

[85] Whiting P, Davies P ,Savović J, Caldwell DM , Churchill R. , *Evidence to inform the development of ROBIS, a new tool to assess the risk of bias in systematic reviews.* [Internet]. 2013 [cited 2016 Oct 20]. Available from: http://www.robis-tool.info)

[86] Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. , *AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews.* J Clin Epidemiol. 2009 Oct;62(10):1013-20. doi: 10.1016/j.jclinepi.2008.10.009.

[87] Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, Davies P, Kleijnen J, Churchill R; ROBIS group. , *ROBIS: a new tool to assess risk of bias in systematic reviews was developed.* J Clin Epidemiol. 2016 Jan;69:225-34. doi: 10.1016/j.jclinepi.2015.06.005. Epub 2015 Jun 16.

[88] Naylor CD. , *Meta-analysis and the metaepidemiology of clinical research.* BMJ. 1997 Sep 13;315(7109):617-9.

[89] Sterne JA, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. , *Statistical methods for assessing the in uence of study characteristics on treatment effects in 'meta-epidemiological' research.* Stat Med. 2002 Jun 15;21(11):1513-24.

[90] Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA., *Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study.* BMJ. 2008 Mar 15;336(7644):601-5. doi: 10.1136/bmj.39465.451748.AD. Epub 2008 Mar 3.

[91] Trinquart L, Dechartres A, Ravaud P., *Commentary: Meta-epidemiology, meta-meta-epidemiology or network meta-epidemiology?* Int J Epidemiol. 2013 Aug;42(4):1131-3. doi: 10.1093/ije/dyt137. Epub 2013 Jul 25.

[92] Dechartres A, Trinquart L, Faber T, Ravaud P., *Empirical evaluation of which trial characteristics are associated with treatment effect estimates.* J Clin Epidemiol. 2016 Sep;77:24-37. doi: 10.1016/j.jclinepi.2016.04.005. Epub 2016 Apr 29. Review.

[93] Valdes AM, Arden NK, Tamm A, Kisand K, Doherty S, Pola E, Cooper C, Tamm A, Muir KR, Kerna I, Hart D, O'Neil F, Zhang W, Spector TD, Maciewicz RA, Doherty M., *A meta-analysis of interleukin-6 promoter polymorphisms on risk of hip and knee osteoarthritis.* Osteoarthritis Cartilage. 2010 May;18(5):699-704. doi: 10.1016/j.joca.2009.12.012. Epub 2010 Feb 6.

[94] Tzoulaki I,Siontis KC, Ioannidis JP., *Prognostic effect size of cardio- vascular biomarkers in datasets from observational studies versus randomised trials: meta-epidemiology study* BMJ 2011; 343

[95] Dechartres A, Trinquart L, Boutron I, Ravaud P. , *Influence of trial sample size on treatment effect estimates: meta-epidemiological study.* BMJ. 2013 Apr 24;346:f2304. doi: 10.1136/bmj.f2304. Review.

[96] Giraudeau B, Higgins JP,Tavernier E, Trinquart L. , *Sample size calculation for meta-epidemiological studies.* Stat Med. 2016 Jan 30;35(2):239-50. doi: 10.1002/sim.6627. Epub 2015 Aug 19.

[97] Zhang Z. , *Meta-epidemiological study: a step-by-step approach by using R.* J Evid Based Med. 2016 Feb 9. doi: 10.1111/jebm.12191. [Epub ahead of print]

[98] Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP., *Evaluating the quality of evidence from a network meta-analysis.* PLoS One. 2014 Jul 3;9(7):e99682. doi: 10.1371/journal.pone.0099682. eCollection 2014.

[99] Salanti G, Marinho V, Higgins JP., *A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered.* J Clin Epidemiol. 2009 Aug;62(8):857-64. doi: 10.1016/j.jclinepi.2008.10.001. Epub 2009 Jan 20.

[100] Chaimani A,Higgins JP, Mavridis D,Spyridonos P,Salanti G., *Graphical tools for network meta-analysis in STATA* PLoS One. 2013 Oct 3;8(10):e76654. doi: 10.1371/journal.pone.0076654. eCollection 2013.

[101] Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A., *GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology.* J Clin Epidemiol. 2011 Apr;64(4):380-2. doi: 10.1016/j.jclinepi.2010.09.011. Epub 2010 Dec 24.

[102] Papageorgiou SN., *Overview provides insights on the current status and future of meta-epidemiology.* J Clin Epidemiol. 2016 Sep;77:11-12. doi: 10.1016/j.jclinepi.2016.05.001. Epub 2016 May 14.

[103] Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, Ramsay T, Bai A, Shukla VK, Grimshaw JM., *External validation of a measurement tool to assess systematic reviews (AMSTAR).* PLoS One. 2007 Dec 26;2(12):e1350.

[104] Berkman ND, Lohr KN, Morgan LC, Kuo TM, Morton SC., *Interrater reliability of grading strength of evidence varies with the complexity of the evidence in systematic reviews.* J Clin Epidemiol. 2013 Oct;66(10):1105-1117.e1. doi: 10.1016/j.jclinepi.2013.06.002.

[105] Brito JP, Tsapas A, Griebeler ML, Wang Z, Prutsky GJ, Domecq JP, Murad MH, Montori VM., Systematic reviews supporting practice guideline recommendations lack protection against bias. J Clin Epidemiol. 2013 Jun;66(6):633-8. doi: 10.1016/j.jclinepi.2013.01.008. Epub 2013 Mar 16. Review.

[106] Faggion Jr CM , Listl S , Giannakopoulos NN., The methodological quality of systematic reviews of animal studies in dentistry. Vet J. 2012 May;192(2):140-7. doi: 10.1016/j.tvjl.2011.08.006. Epub 2011 Sep 15. Review.

[107] Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, Maida CA., From systematic reviews to clinical recommendations for evidence-based health care: validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for grading of clinical relevance. Open Dent J. 2010 Jul 16;4:84-91. doi: 10.2174/1874210601004020084.

[108] Pieper D, Buechter RB, Li L, Prediger B, Eikermann M., Systematic review found AMSTAR, but notR(evised)-AMSTAR, to have good measurement properties J Clin Epidemiol. 2015 May;68(5):574-83. doi: 10.1016/j.jclinepi.2014.12.009. Epub 2014 Dec 30. Review.

[109] Burda BU, Norris SL, Holmer HK, Ogden LA, Smith ME., Quality varies across clinical practice guidelines for mammography screening in women aged 40-49 years as assessed by AGREE and AMSTAR instruments. J Clin Epidemiol. 2011 Sep;64(9):968-76. doi: 10.1016/j.jclinepi.2010.12.005. Epub 2011 Mar 21. Review.

[110] Monasta L, Batty GD, Cattaneo A, Lutje V, Ronfani L, Van Lenthe FJ, Brug J. , Early-life determinants of overweight and obesity: a review of systematic reviews. Obes Rev. 2010 Oct;11(10):695-708. doi: 10.1111/j.1467-789X.2010.00735.x. Review.

[111] Needleman I, Clarkson J, Worthington H., A practitioner's guide to developing critical appraisal skills: reviews of research. J Am Dent Assoc. 2013 May;144(5):527-30.

[112] Woolf SH,Grol R, Hutchinson A ,Eccles M , Grimshaw J. , Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ. 1999 Feb 20;318(7182):527-30. Review.

[113] Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Rada G, Rosenbaum S, Morelli A, Guyatt GH, Oxman AD; GRADE Working Group., GRADEWorking Group.GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016 Jun 28;353:i2016. doi: 10.1136/bmj.i2016.

[114] Phillips B, Ball C ,Sackett D, Badenoch D , Straus S , Haynes B , Dawes M., Evidence-based medicine levels of evidence.. Oxford Centre for Evidence-Based Medicine.2001.

[115] Phillips B, Ball C ,Sackett D, Badenoch D , Straus S , Haynes B , Dawes M., A new system for grading recommendations in evidence based guidelines. BMJ. 2001 Aug 11;323(7308):334-6.

[116] Guyatt G, Schünemann HJ, Cook D, Jaeschke R, Pauker S., Applying the grades of recommendation for antithrombotic and thrombolytic therapy: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy Chest. Chest. 2004 Sep;126(3 Suppl):179S-187S.

[117] Upshur RE ., Are all evidence-based practices alike? Problems in the ranking of evidence. CMAJ. 2003 Sep 30;169(7):672-3.

[118] Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, Liberati A, O'Connell D, Oxman AD, Phillips B, Schünemann H, Edejer TT, Vist GE, Williams JW Jr; GRADE Working Group., Systems for grading the quality of evidence and the strength of recommendations I. Critical appraisal of existing approaches. The GRADE Working Group . BMC Health Serv Res. 2004 Dec 22;4(1):38.

[119] Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, Hill S, Jaeschke R, Liberati A, Magrini N, Mason J, O'Connell D, Oxman AD,

Phillips B, Schünemann H, Edejer TT, Vist GE, Williams JW Jr; GRADE Working Group., Systems for grading the quality of evidence and the strength of recommendations I. Critical appraisal of existing approaches. The GRADE Working Group . BMC Health Serv Res. 2005 Mar 23;5(1):25.

[120] Aguayo-Albasini JL, Flores-Pastor B, Soria-Aledo V.., GRADE system:classification of quality of evidence and strength of recommendation. Cir Esp. 2014 Feb;92(2):82-8. doi: 10.1016/j.ciresp.2013.08.002. Epub 2013 Dec 20. Spanish.

[121] Guyatt GH, Norris SL, Schulman S, Hirsh J, Eckman MH, Akl EA, Crowther M, Vandvik PO, Eikelboom JW, McDonagh MS, Lewis SZ, Gutterman DD, Cook DJ, Schünemann HJ. GRADE system:classification of quality of evidence and strength of recommendation. Chest. 2012 Feb;141(2 Suppl):53S-70S. doi: 10.1378/chest.11-2288.

[122] Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Vandvik PO, Meerpohl J, Guyatt GH, Schünemann HJ; GRADE Working Group. [GRADE Evidence to Decision(EtD)frameworks: a systematic and transparent approach to making well informed health care choices.2:Clinical practice guidelines] BMJ. 2016 Jun 30;353:i2089. doi: 10.1136/bmj.i2089.

[123] Wallentin L, Yusuf S, Ezekowitz MD, Alings M, Flather M, Franzosi MG, Pais P, Dans A, Eikelboom J, Oldgren J, Pogue J, Reilly PA, Yang S, Connolly SJ; RELY investigators. [GRADE Evidence to Decision(EtD)frameworks: a systematic and transparent approach to making well informed health care choices.2:Clinical practice guidelines] Lancet. 2010 Sep 18;376(9745):975-83.

[124] Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, Atkins D, Kunz R, Montori V, Jaeschke R, Rind D, Dahm P, Akl EA, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Schünemann HJ. [GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes.] J Clin Epidemiol. 2013 Feb;66(2):151-7. doi: 10.1016/j.jclinepi.2012.01.006. Epub 2012 Apr 27. Review.

[125] Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J, Sisk J, Ruiz F, Hill S, Guyatt GH, Jaeschke R, Helfand M, Harbour R, Davoli M, Amato L, Liberati A, Schünemann HJ.] GRADE guidelines: 10. Considering resource use

and rating the quality of economic evidence. J Clin Epidemiol. 2013 Feb;66(2):140-50. doi: 10.1016/j.jclinepi.2012.04.012. Epub 2012 Aug 3. Review.

[126] Beauchamp TL and Childress JF.] Principles of biomedical ethics.7th ed. Oxford University Press. 2013.Oxford.

[127] Flottorp SA, Oxman AD, Krause J, Musila NR, Wensing M, Godycki-Cwirko M, Baker R, Eccles MP. A checklist for identifying determinants of practice: a systematic review and synthesis of frameworks and taxonomies of factors that prevent or enable improvements in healthcare professional practice. Implement Sci. 2013 Mar 23;8:35. doi: 10.1186/1748-5908-8-35. Review.

[128] Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, Nasser M, Meerpohl J, Post PN, Kunz R, Brozek J, Vist G, Rind D, Akl EA, Schünemann HJ., GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol. 2013 Jul;66(7):719-25. doi: 10.1016/j.jclinepi.2012.03.013. Epub 2013 Jan 9.

[129] Fretheim A, Oxman AD, Lavis JN, Lewin S., SUPPORT Tools for Evidence-informed Policymaking in health 18: Planning monitoring and evaluation of policies. Health Res Policy Syst. 2009 Dec 16;7 Suppl 1:S18. doi: 10.1186/1478-4505-7-S1-S18.

[130] Munafo M, Nosek BA, Bishop D, Button KS, Chambers CD, Percie du Sert N, Simonsohn U, Wagenmakers EJ, Ware and John P. Ioannidis A., A manifesto for reproducible science. Nature Human Behaviour 1, Article number: 0021 (2017) doi:10.1038/s41562-016-0021

Escalas y sistemas de conducción y evaluación

Tabla 1: Criterios evaluación del riesgo de sesgo.

1. **No se aconseja emplear escalas de calidad.**	1.1.-No son una forma adecuada porque tienden a combinar evaluaciones de aspectos de la calidad de la presentación de informes con los de la conducción y asignar pesos a los ítems que son difíciles de justificar. 1.2.-Tanto las consideraciones teóricas como la evidencia empírica sugieren que las asociaciones de diferentes escalas con las estimaciones del efecto de la intervención son inconsistentes e impredecibles.
2. **Centrarse en la validez interna.** 2.2. Es importante separarla de la validez externa y la precisión.	2.1. La validez interna es la medida en que está libre de sesgos.
3. **Evaluar el riesgo de sesgo de los resultados de los ensayos clínicos no la calidad de los informes o problemas metodológicos no directamente relacionados con el riesgo de sesgo.**	3.1. La calidad de la información afecta la capacidad de evaluación del riesgo de sesgo. 3.2. Algunos aspectos de la conducta del ensayo clínico no están directamente relacionados con el riesgo de sesgo. 3.3. Altos estándares de la calidad metodológica no implican ausencia de riesfo de sesgo.
4. **Las evaluaciones del riesgo de sesgo requieren un juicio.**	4.1. La evaluación de un aspecto particular de la conducta del ensayo clínico requiere tanto el conocimiento de los métodos como un juicio sobre si éstos pueden haber llevado a un riesgo de sesgo. La base para las evaluaciones de sesgo debería ser explícita, registrándose los aspectos en los que se basó el juicio .
5. **Elejir los dominios que se evaluarán en base a consideraciones teóricas y empíricas.**	5.1. Determinados aspectos de la conducta del ensayo clínico están asociados con sesgos. Para otros la evidencia aún no es clara. Finalmente, puede haber algunos específicos del diseño que son relevantes sólo para determinados ensayos clínicos y revisiones sistemáticas.

6. **Enfocar el riesgo de sesgo en los datos tal como se representa en la revisión sistemática en lugar de como se informó originalmente.**	6.1. Algunos artículos pueden notificar los resultados de los ensayos clínicos que se consideran de alto riesgo de sesgo, para lo cual es posible derivar un resultado con bajo riesgo de sesgo.
7. **Comunicar evaluaciones específicas del riesgo de sesgo.**	Algunos aspectos de la conducta del ensayo se aplican al ensayo en su conjunto. Sin embargo, para otros, el riesgo de sesgo es específico de los resultados dentro del ensayo.

Tabla 2: Tipos de sesgo de publicación.

Sesgo de publicación.	*La publicación depende de la naturaleza y dirección de los resultados.*
Sesgo de lapso de tiempo.	*La publicación rápida o demorada depende de la naturaleza de los resultados.*
Sesgo de publicación múltiple (duplicada).	*La publicación múltiple o única de los hallazgos depende de la naturaleza de los resultados .*
Sesgo de ubicación.	*El nivel de acceso a la revista o la indexación depende de la naturaleza y dirección de los resultados.*
Sesgo de citación.	*La citación depende de la naturaleza y dirección de los resultados.*
Sesgo de idioma.	*La publicación en un idioma particular depende de la naturaleza o dirección de los resultados.*
Sesgo de informe de resultado.	*El informe selectivo depende de la naturaleza y/o dirección de los resultados.*

Tabla 3: PRISMA.

Section/topic. (#Item)	Item checklist
TITLE	
Title (#1)	Identify the report as a systematic review, meta-analysis, or both.
ABSTRACT	
Structured summary (#2)	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.
INTRODUCTION	
Rationale (#3)	Describe the rationale for the review in the context of what is already known.
Objectives (#4)	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).
METHODS	
Protocol and registration (#5)	Indicate whether a review protocol exists and if and where it can be accessed (e.g., Web address); and, if available, provide registration information, including registration number.
Eligibility criteria (#6)	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.
Information sources (#7)	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
Search (#8)	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
Study selection (#9)	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
Data collection process (#10)	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data items (#11)	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Risk of bias in individual studies (#12)	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary measures (#13)	State the principal summary measures (e.g., risk ratio, difference in means).

Synthesis of results (#14)	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I2) for each meta-analysis.
Risk of bias across studies (#15)	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional analyses (#16)	Describe methods of additional analyses if done, indicating which were pre-specified.
RESULTS	
Study selection (#17)	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics (#18)	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.
Risk of bias within studies (#19)	Present data on risk of bias of each study and, if available, any outcome level assessment (see item #12).
Results of individual studies (#20)	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of results (#21)	Present results of each meta-analysis done, including confidence intervals and measures of consistency.
Risk of bias across studies (#22)	Present results of any assessment of risk of bias across studies (see item #15).
Additional analysis (#23)	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see item #16]).
DISCUSSION	
Summary of evidence (24)	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).
Limitations (25)	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).
Conclusions (26)	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
FUNDING	
Funding (27)	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

Tabla 4: PRISMA for network meta-analysis (NMA)

Section/topic. (#Item)	Checklist item
TITLE	
Title (#1)	Identify the report as a systematic review incorporating a network meta-analysis (or related form of meta-analysis).
ABSTRACT	
Structured summary (#2)	Provide a structured summary including, as applicable: • Background: main objectives. • Methods: data sources; study eligibility criteria, participants, and interventions; study appraisal; and synthesis methods, such as network meta-analysis. • Results: number of studies and participants identified; summary estimates with corresponding confidence/credible intervals; treatment rankings may also be discussed. Authors may choose to summarize pairwise comparisons against a chosen treatment included in their analyses for brevity. • Discussion/Conclusions: limitations; conclusions and implications of findings. • Other: primary source of funding; systematic review registration number with registry name
INTRODUCCTION	
Rationale (#3)	Describe the rationale for the review in the context of what is already known, including mention of why a network meta-analysis has been conducted.
Objectives (#4)	Provide an explicit statement of questions being addressed, with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).
METHODS	
Protocol and registration (#5)	Indicate whether a review protocol exists and if and where it can be accessed (e.g., Web address); and, if available, provide registration information, including registration number.
Eligibility criteria (#6)	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. Clearly describe eligible treatments included in the treatment network, and note whether any have been clustered or merged into the same node (with justification).
Information sources (#7)	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.

Search (#8)	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
Study selection (#9)	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
Data collection process (#10)	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data items (#11)	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Geometry of the network (**#S1**)	Describe methods used to explore the geometry of the treatment network under study and potential biases related to it. This should include how the evidence base has been graphically summarized for presentation, and what characteristics were compiled and used to describe the evidence base to readers.
Risk of bias whitin individual studies (#12)	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary measures (#13)	State the principal summary measures (e.g., risk ratio, difference in means). Also describe the use of additional summary measures assessed, such as treatment rankings and surface under the cumulative ranking curve (SUCRA) values, as well as modified approaches used to present summary findings from meta-analyses.
Planned methods of analysis (#14)	Describe the methods of handling data and combining results of studies for each network meta-analysis. This should include, but not be limited to: • Handling of multi-arm trials; • Selection of variance structure; • Selection of prior distributions in Bayesian analyses; and • Assessment of model fit.
Assessment of Inconsistency (**#S2**)	Describe the statistical methods used to evaluate the agreement of direct and indirect evidence in the treatment network(s) studied. Describe efforts taken to address its presence when found.
Risk of bias across studies (15)	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional analyses (#16)	Describe methods of additional analyses if done, indicating which were pre-specified. This may include, but not be limited to, the following: • Sensitivity or subgroup analyses; • Meta-regression analyses; • Alternative formulations of the treatment network; and Use of alternative prior distributions for Bayesian analyses (if applicable).
RESULTS	
Study selection (#17)	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.

Presentation of network structure (**#S3**)	Provide a network graph of the included studies to enable visualization of the geometry of the treatment network.
Summary of network geometry (**#S4**)	Provide a brief overview of characteristics of the treatment network. This may include commentary on the abundance of trials and randomized patients for the different interventions and pairwise comparisons in the network, gaps of evidence in the treatment network, and potential biases reflected by the network structure.
Study characteristics (#18)	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.
Risk of bias within studies (#19)	Present data on risk of bias of each study and, if available, any outcome level assessment .
Results of individual studies (#20)	For all outcomes considered (benefits or harms), present, for each study: 1) simple summary data for each intervention group, and 2) effect estimates and confidence intervals. Modified approaches may be needed to deal with information from larger networks.
Synthesis of results (#21)	Present results of each meta-analysis done, including confidence/credible intervals. In larger networks, authors may focus on comparisons versus a particular comparator (e.g. placebo or standard care), with full findings presented in an appendix. League tables and forest plots may be considered to summarize pairwise comparisons. If additional summary measures were explored (such as treatment rankings), these should also be presented.
Exploration for inconsistency (**#S5**)	Describe results from investigations of inconsistency. This may include such information as measures of model fit to compare consistency and inconsistency models, P values from statistical tests, or summary of inconsistency estimates from different parts of the treatment network.
Risk of bias across studies (#22)	Present results of any assessment of risk of bias across studies for the evidence base being studied.
Additional analysis (#23)	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression analyses, alternative network geometries studied, alternative choice of prior distributions for Bayesian analyses, and so forth). .
DISCUSSION	
Summary of evidence (#24)	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).
Limitations (#25)	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias). Comment on the validity of the assumptions, such as transitivity and consistency. Comment on any concerns regarding network geometry (e.g., avoidance of certain comparisons).

Conclusions (#26)	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
FUNDING	
Funding (#27)	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. This should also include information regarding whether funding has been received from manufacturers of treatments in the network and/or whether some of the authors are content experts with professional conflicts of interest that could affect use of treatments in the network.

Tabla 5: PRISMA for Abstracts (PRISMA-A).

Section/topic. (#Item)	Checklist item
TITLE	
Title (#1)	Identify the report as a systematic review, meta-analysis, or both.
BACKGROUND	
Objectives (#2)	The research question including components such as participants, interventions, comparators, and outcomes.
METHODS	
Eligibility criteria (#3)	Study and report characteristics used as criteria for inclusion.
Information sources: (#4)	Key databases searched and search dates.
Risk of bias: (#5)	Methods of assessing risk of bias.
RESULTS	
Included studies (#6)	Number and type of included studies and participants and relevant characteristics of studies.
Synthesis of results (#7)	Results for main outcomes (benefits and harms), preferably indicating the number of studies and participants for each. If meta-analysis was done, include summary measures and confidence intervals.
Description of the effect (#8)	Direction of the effect (i.e. which group is favoured) and size of the effect in terms meaningful to clinicians and patients.
DISCUSSION	
Strengths and Limitations of evidence: (9)	Brief summary of strengths and limitations of evidence (e.g. inconsistency, imprecision, indirectness, or risk of bias, other supporting or conflicting evidence)
Interpretation(10)	General interpretation of the results and important implications
OTHER	
Funding (11)	Primary source of funding for the review.
Registration (12))	Registration number and registry name.

Tabla 6: Características de los estudios meta epidemiológicos, meta-metaepidemiológicos y metaepidemiológicos en red.

	Metaepidemiología	Meta-metaepidemiología	Meta-epidemiología en red
Fuentes de datos.	Meta-análisis de ensayos clínicos	Estudios de meta-epidemiológicos , harmonizados, sin overlap entre meta-análisis	Meta-análisis en red.
Restricciones.	Meta-análisis con ensayos clínicos con y sin el factor de estudio	Diferentes estudios de metaepidemiológicos deben investigar varios conjuntos de factores de riesgo, potencialmente evaluados con diferentes métodos	Meta-análisis en red deben incluir más ensayos que intervenciones.
Evaluación de los factores de riesgo relacionados con el nivel de prueba	La reevaluación de los informes de ensayos individuales o la dependencia de la evaluación de cada meta-análisis seleccionada	Evaluación de cada estudio meta-epidemiologíco	Reevaluación de informes de ensayos individuales o dependencia en la evaluación de cada red seleccionada.
Asunción respecto a la dirección del sesgo.	En las comparaciones activo-inactivo, no se espera que un factor de riesgo favorezca al comparador inactivo. En las comparaciones activo-activo, se requiere una suposición con respecto a la dirección del sesgo.	En las comparaciones activo-inactivo, no se espera que un factor de riesgo favorezca al comparador inactivo. En las comparaciones activo-activo, se requiere una suposición con respecto a la dirección del sesgo.	En las redes en forma de estrella, se espera que un factor de riesgo no favorezca al comparador común. En redes con bucles cerrados, es necesario un supuesto con respecto a la dirección del sesgo.
Estimación del impacto de los factores de riesgo en las estimaciones del efecto de la intervención.	Las estimaciones de efectos se comparan entre los ensayos con y sin el factor de riesgo dentro de cada MA; el impacto significativo del factor de riesgo se estima en todas las AM.	Las estimaciones de efectos se comparan entre los ensayos con y sin el factor de riesgo dentro de cada MA; el impacto significativo del factor de riesgo se estima en todas las AM.	Las estimaciones de efectos se comparan entre los ensayos con y sin el factor de riesgo dentro de cada red; el impacto medio del factor de riesgo se estima en todas las redes.

Asunción respecto a la intercambiabilidad del impacto de los factores de riesgo en las estimaciones del efecto de la intervención.	Entre ensayos dentro de meta-análisis, y entre meta-análisis	Entre ensayos dentro de meta-análisis, y entre meta-análisis.	Entre los ensayos dentro de las redes y entre las comparaciones de la red.

Tabla 7: AMSTAR.

Items	Justificación y Razonamientos de apoyo.
1. *Was an 'a priori' design provided?*	The research question and inclusion criteria should be established before the conduct of the review. **Note:** Need to refer to a protocol, ethics approval, or pre-determined/a priori published research objectives to score a "yes."
2. *Was there duplicate study selection and data extraction?*	There should be at least two independent data extractors and a consensus procedure for disagreements should be in place. **Note:** Two people do study selection, two people do data extraction, consensus process or one person checks the other's work.
3. *Was a comprehensive literature search performed?*	At least two electronic sources should be searched. The report must include years and databases used (e.g., Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found. **Note:** If at least 2 sources + one supplementary strategy used, select "yes" (Cochrane register/Central counts as 2 sources; a grey literature search counts as supplementary).
4. *Was the status of publication (i.e. grey literature) used as an inclusion criterion?*	The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc. **Note:** If review indicates that there was a search for "grey literature" or "unpublished literature," indicate "yes." SINGLE database, dissertations, conference proceedings, and trial registries are all considered grey for this purpose. If searching a source that contains both grey and non-grey, must specify that they were searching for grey/unpublished lit.
5. *Was a list of studies (included and excluded) provided?*	A list of included and excluded studies should be provided. **Note:** Acceptable if the excluded studies are referenced. If there is an electronic link to the list but the link is dead, select "no."
6. *Were the characteristics of the included studies provided?*	In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported. **Note:** Acceptable if not in table format as long as they are described as above.

7. *Was the scientific quality of the included studies assessed and documented?*	'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant. **Note:** Can include use of a quality scoring tool or checklist, e.g., Jadad scale, risk of bias, sensitivity analysis, etc., or a description of quality items, with some kind of result for EACH study ("low" or "high" is fine, as long as it is clear which studies scored "low" and which scored "high"; a summary score/range for all studies is not acceptable).
8. *Was the scientific quality of the included studies used appropriately in formulating conclusions?*	The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations. **Note:** Might say something such as "the results should be interpreted with caution due to poor quality of included studies." Cannot score "yes" for this question if scored "no" for question 7.
9. *Were the methods used to combine the findings of studies appropriate?*	For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e., Chi-squared test for homogeneity, I2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?). **Note:** Indicate "yes" if they mention or describe heterogeneity, i.e., if they explain that they cannot pool because of heterogeneity/variability between interventions
10. *Was the likelihood of publication bias assessed?*	An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test, Hedges and Olkin's method). **Note:** If no test values or funnel plot included, score "no". Score "yes" if mentions that publication bias could not be assessed because there were fewer than 10 included studies.
11. *Was the conflict of interest included?*	Potential sources of support should be clearly acknowledged in both the systematic review and the included studies. **Note:** To get a "yes," must indicate source of funding or support for the systematic review AND for each of the included studies.

Tabla 8: Fase II de ROBIS. Dominio 1.

Preguntas de señalización	Razonamiento
1.1. ¿Se adhirió la revisión sistemática a los objetivos predefinidos y a los criterios de elección?.	La respuesta a esta pregunta viene dada por la presencia de un protocolo o por el juicio del revisor sobre lo notificado en el informe del estudio. Se recomienda responder: • “Sí”: Cuando la información está disponible en un protocolo. • "Probablemente sí": si existe evidencia de que los objetivos y los criterios de elección fueron especificados *a priori*. • "Probablemente no": si sólo están disponibles post hoc. • “No": si faltan todos los detalles sobre los objetivos y criterios de elección.
1.2. ¿Fueron los criterios de elección apropiados para la pregunta de revisión?	Para responder a esta cuestión el evaluador requiere de conocimientos específicos del tema de revisión y los criterios deben estar suficientemente detallados para permitir que se evalúe si los estudios incluidos son los apropiados para responder la pregunta de investigación.
1.3. ¿Existe ambigüedad en los criterios de elección?	Los criterios deben estar suficientemente detallados para que pueda replicarse. Si surgen preguntas específicas sobre los criterios de elegibilidad que no pueden ser respondidas con la información disponible deben marcarse los juicios "no" o "Probablemente no".
1.4. ¿Fueron apropiadas las restricciones de los criterios de elección basadas en las características de los estudios?	Cuando no se informan explícitamente restricciones sobre las características del estudio o cuando se dispone de información suficiente, y el evaluador está razonablemente satisfecho de que las restricciones son apropiadas puede responderse a esta pregunta "sí o" Probablemente sí ". Sin embargo, cuando las restricciones no están justificadas o la información no está disponible debe responderse : "Probablemente no" o "no".
1.5. ¿Fueron apropiadas las restricciones basadas en las fuentes de los estudios?	Este apartado se refiere al estado o formato de la publicación, el idioma y la disponibilidad de los datos. Cualquier restricción debe ser claramente expuesta y acompañada de una sólida justificación. Cuando no existen restricciones o estas se detallan y son apropiadas la pregunta debe ser contestada "sí" o “probablemente sí”. En caso contrario se valora como “no” o “probablemente no”. Sino existe información se valora como “ incierta”.

Tabla 9: Fase II de ROBIS. Dominio 2.

Preguntas de señalización	Razonamiento
2.1. ¿La búsqueda incluyó un rango apropiado de bases de datos y fuentes electrónicas para informes publicadas y no publicados?.	Varía según el tema de la revisión sistemática. Como mínimo, debe buscarse en MEDLINE y EMBASE, además de búsquedas de informes de conferencias y de registros de investigación.
2.2. ¿Se utilizaron otros métodos adicionales a la búsqueda en bases de datos electrónicas para la identificación de informes?	Deben realizarse métodos adicionales tales como: búsquedas de citas, contacto con expertos, búsqueda manual, etc. Las respuestas a las preguntas de señalización se realizan en base a la consideración del grado de cumplimentación de la búsqueda.
2.3. ¿Fueron los términos y la estructura de la estrategia de la búsqueda adecuados para recuperar la mayor cantidad de estudios?	Se requiere una estrategia de búsqueda completa que muestre todos los términos utilizados para poder ser replicada y juzgada esta pregunta. Los evaluadores deben considerar si la estrategia de búsqueda incluyó una gama adecuada de términos, la combinación apropiada de los mismos y las palabras del título y del resumen utilizadas además del empleo de los filtros adecuados. Puede encontrarse orientación sobre las estrategias de búsqueda. • Si todo ello se realiza de forma adecuada puede evaluarse como “sí”. • Si sólo se proporcionan detalles limitados e incompletos,puede justificarse un "Probablemente sí" o "Probablemente no". • “No” cuando se haya realizado de forma errónea.
2.4. ¿Se realizaron correctamente las restricciones basadas en la fecha, el formato de la publicación y el idioma?	Se requiere información sobre los tres componentes de esta pregunta para juzgarla. • Si no se aplican restricciones, debe responderse “sí”. • La restricción basada en el lenguaje o formato de publicación rara vez es apropiada; si se aplicaron debe contestarse como "No". • Las restricciones en la fecha pueden ser apropiadas pero deben ser apoyadas por una justificación para que esta pregunta sea contestada como "sí".
2.5. ¿Se realizaron los esfuerzos adecuados para la minimización de errores en la selección de los estudios?	Esta pregunta comprende tanto la selección de títulos y resúmenes como de evaluación del texto completo de los estudios. Para una respuesta de "Sí” el proceso debe haber sido realizado por al menos dos revisores, idealmente de forma independientemente, o con uno realizando la evaluación y el segundo comprobando la decisión.

Tabla 10: Fase II de ROBIS. Dominio 3.

Preguntas de señalización	Razonamiento
3.1. ¿Se realizaron los esfuerzos para disminuir el error en la recogida de los datos?	El proceso de recogida de datos debe ser realizado por dos revisores con formularios estructurados que hayan sido pilotados. Lo ideal es que se realice de forma independiente. Sin embargo, la extracción por un revisor y la verificación detallada por un segundo revisor también es aceptable.
3.2. Se recopilaron las suficientes características de los estudios para que los revisores y los lectores puedan interpretar los datos?	La información sobre las características del estudio permite una investigación de la heterogeneidad y la consideración de la aplicabilidad de los resultados. Puede estar disponible a partir de los cuadros de los estudios incluidos o resumirse en el texto de los resultados. Esta pregunta es difícil de juzgar por las restricciones de espacio del informe. En muchas ocasiones es necesario que los evaluadores accedan a recursos adicionales como los apéndices-web.
3.3. ¿Se recogieron todos los resultados relevantes para la síntesis de los datos?	Los autores deben informar a priori qué datos se requieren para la síntesis y en qué formato. Es muy raro que todos los estudios primarios incluidos informen los datos en el formato apropiado. Para responder "sí" a esta pregunta,debe incluirse información detallada, en la sección de métodos para describir cómo se obtuvieron los datos de resultados que no se informaron en el formato requerido para la síntesis: mediante estimación, transformación o contactando con los autores para obtener información adicional.
3.4. ¿Se evaluó adecuadamente el riesgo de sesgo de los estudios primarios?	Si el riesgo de sesgo no fue evaluado formalmente esta pregunta debe ser contestada como "no". Si se llevó a cabo una evaluación formal, los evaluadores necesitarán utilizar su criterio para determinar si es apropiado: Si se utilizó una herramienta publicada y validada para el diseño,esta pregunta debe contestarse como sí. Si la revisión sólo enumera las preguntas evaluadas, utiliza una herramienta no publicada o que ya no se recomienda, el evaluador necesita juzgar si los criterios revisados por la herramienta fueron suficientes para identificar fuentes potenciales de sesgo en los estudios primarios. Por ejemplo, la puntuación de Jadad no incluye el ocultamiento de la asignación. Para responder "sí" las revisiones sistemáticas que han utilizado la escala de Jadad deberían evaluar el ocultamiento de la asignación.
3.5. ¿Se realizaron los esfuerzos para minimizar el sesgo en la evaluación de los estudios primarios?	La evaluación del riesgo de sesgo debe ser llevada a cabo por al menos dos revisores. Idealmente de forma independientemente, pero la evaluación por un revisor y el control por un segundo también es aceptable.

Tabla 11: Fase II de ROBIS. Dominio 4.

Preguntas de señalización	Razonamiento
4.1. ¿La síntesis incluyó todos los estudios que debía?	Generalmente, la síntesis debe incluir todos los estudios que tienen datos pertinentes. Los resultados de estudios individuales pueden faltar en la síntesis, además de por un sesgo de publicación, porque: • Los revisores no tienen acceso a los resultados específicos: los estudios incluidos no informaron los resultados por su falta de significación estadística o por limitaciones de espacio. • Los revisores no han podido recopilar o procesar los datos disponibles: los estudios pueden haber sido omitidos por error o porque los revisores desconocen los cálculos estadísticos que permitirían su inclusión. Esto sería problemático si los estudios omitidos tuvieran resultados sistemáticamente diferentes de aquello que sí se incluyeron. • Los revisores han excluido deliberadamente los resultados. Un ejemplo sería la exclusión de estudios basada sólo en consideraciones estadísticas.
4.2. ¿Se siguieron todos los análisis predefinidos o se explicaron las desviaciones?	El propósito de esta pregunta es identificar los sesgos introducidos mediante la selección de análisis y métodos de análisis porque los resultados que no les gustan son suprimidos o reemplazados. Para responder “sí” la revisión debería disponer un protocolo publicado o accesible. Si hay una indicación de que se predefinieron los análisis, el evaluador podría responder "probablemente sí". En la ausencia explícita de un protocolo a priori debe responderse "no". Si no se hace referencia a la existencia o ausencia de un protocolo se recomienda responder: "ninguna información".
4.3. ¿Fue adecuada la síntesis dada la naturaleza y similitud de las preguntas de investigación, el diseño de estudio y los resultados a través de los trabajos incluidos?	Es necesario analizar la heterogeneidad clínica y estadística y evaluar si el resultado final es significativo para la toma de decisiones. El juicio sobre síntesis cuantitativa se refiere a su pertinencia y métodos estadísticos utilizados. La valoración del enfoque narrativo se refiere,a su pertinencia y a si el método empleado es el apropiado.
4.4. ¿Fue la variabilidad de los resultados entre los estudios mínima o recogida en la síntesis?	Si se ignora una heterogeneidad sustancial en un meta-análisis se puede dar lugar a conclusiones engañosas y/o una falsa precisión. Si un meta-análisis de efectos fijos se utiliza en presencia de heterogeneidad es importante que los revisores notifiquen que el análisis ignora la heterogeneidad. En caso contrario debería juzgarse como "No". Si se ha utilizado adecuadamente un modelo de efectos aleatorios para la heterogeneidad y/o análisis de subgrupos o meta-regresión, podría responderse "Sí". Si se realizó una síntesis narrativa sobre la base de la combinación fue inapropiada debido a la heterogeneidad clínica esta pregunta se debe responder "Sí”.

4.5. ¿Fueron los hallazgos suficientemente,robustos demostrados a través de gráficos de embudo o de análisis de sensibilidad?	Pueden emplearse diagramas de embudo para examinar las relaciones entre el tamaño del efecto y,del estudio. Debe estudiarse la simetría y enfocar el tipo de metaánálisis en función de ella. En caso de asimetría debe realizarse análisis de sensibilidad. Si hay muy pocos estudios o son muy heterogéneos, puede estar claro que los resultados no son robustos, incluso si los revisores no realizaron análisis de sensibilidad. Si se realizó una síntesis narrativa se debe,considerar si distintos enfoques para resumir los estudios podrían haber dado lugar a diferentes conclusiones.
4.6. ¿Los sesgos en los estudios primarios fueron mínimos o se abordaron en la síntesis?	Si se ha evaluado el riesgo de sesgo, debe considerarse si se ha tenido en cuenta dicha evaluación en las conclusiones y si se ha actuado conforme a los resultados. Debería juzgarse “sí” si los estudios han recibido calificaciíon de "bajo riesgo de sesgo" o se ha empleado anílisis de sensibilidad para los de alto riesgo. Debe juzgarse "No" si se han encontrado sesgos en los estudios que han sido ignorados por la revisores o no los han incorporado a los hallazgos y conclusiones. Tambiíen si los sesgos se abordan síolo como parte de la discusión de los resultados.

Tabla 12: Fase III de ROBIS.

Preguntas de señalización	Razonamiento
1. ¿La interpretación de los hallazgos aborda todas las preocupaciones identificadas en los dominios 1 a 4.?	Esta pregunta se refiere a la evaluación de las preocupaciones de sesgos encontradas en la fase 2. • Si no se identificaron problemas o las limitaciones se consideraron adecuadamente en las conclusiones se puede calificar como "sí". • Si uno más dominios se calificaron como de “alto riesgo” o “riesgo poco claro” es necesario considerar si se ha tenido en cuenta en las conclusiones y en la interpretación de los hallazgos. Si esto no es así el juicio debería ser “no” o “probablemente no”.
2. ¿Se consideró adecuadamente la relevancia de los estudios encontrados para responder la pregunta de investigación?	Al interpretar los resultados debe evaluarse la pertinencia (aplicabilidad / validez externa) de los estudios incluídos para responder la pregunta de investigación. Cuando no son directamente aplicables y no se considera al interpretar los resultados puede existir riesgo de sesgo. En algunas revisiones sistemáticas se puede considerar la relevancia de los estudios incluidos como parte de la evaluación formal de la calidad, como en las de precisión diagnóstica. Otras pueden emplear la discusión para valorar de la pertinencia de los estudios. Ambos enfoques pueden ser apropiados si en las conclusiones se presenta una reflexión de la relevancia de los estudios incluidos.
3.-¿La RS evita enfatizar los resultados en base a su significación estadística?	Cuando la revisión presenta varios análisis es importante hacer un balance de todos ellos evitando destacar los resultados por su significación estadística.

Tabla 12: Recomendaciones GRADE.

Significado	Fuertes	Débiles
Probablemente apropiadas	Se cumplen todas las condiciones	Se cumple alguna de las condiciones
Calidad de la evidencia	Alta o moderada (o baja o muy baja en circunstancias excepcionales)	Baja o muy baja
Balance riesgos beneficios	Una alternativa es claramente superior	El balance de beneficios y riesgos es cercano
Valores y preferencias de los pacientes	Todos o casi todos los pacientes informados toman la misma decisión	Existe variabilidad e incertidumbre respecto a lo que decidirán pacientes informados.
Consideraciones de recursos	El coste de la intervención está plenamente justificado	El costo de la intervención pudiera no estar justificado en algunas circunstancias.

Lista de figuras

Lista de tablas

Índice de abreviaturas

AA: Acontecimientos adversos.
AAD: American Academy of Dermatology.
AAP: An\'alisis por protocolo.
ACP: An\'alisis por componente principal.
AETSA: Agencia de Evaluaci\'on de Tecnolog\'ias Sanitarias.
AMSTAR: A Measurement Tool to Assess the Methodological Quality of Systematic Revie
AVC: año de vida ajustado por calidad.
BAD: British Academy of Dermatology.
BSA: Body Surface Area.
BMG:Cochrane Bias Methods Group
CENTRAL: Registro Central Cochrane de Ensayos Controlados.
CSG: Cochrane Skin Group.
DARE: Base de Datos de Res\'umenes de Revisiones de Efectos.
DLQI: Dermatology Quality of Life Index.
ECA: ensayo cl\'inico aleatorizado.
GRADE: Grading of Recommendations Assessment, Development and Evaluation.
GPC: Gu\'ias de pr\'actica cl\'inica.
HR: Hazard ratio.
HTA: Base de Datos de Evaluaci\'on de Tecnolog\'ia en Salud.
IGA: Investigator´s Global Assessment.
IL: Interleucina.
ITT: An\'alisis por intenci\'on de tratar.
JBI: Joanna Brigss Institute.
MA: Metaan\'alisis.
NMA: Metaan\'alisis en red.
ME: Metaepidemiolog\'ia.
MECIR: Methodological Expectations of Cochrane Intervention Reviews.
MER: Metaepidemiolog\'ia en red.

MME: Meta-metaepidemiolog\'ia.
OCEBM: Oxford Centre for Evidence-Based Medicine.
OMS: Organizaci\'on mundial de la salud.
OQAQ: cuestionario de evaluaci\'on de la calidad
PASI: Psoriasis Area and Severity Index.
PGA: Physician´s Global Assessment.
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
PRISMA-A: PRISMA for Abstract.
PRISMA-NMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Ne
NNT: N\'umero necesario a tratar.
OR: Odds ratio.
PROSPERO: The International Prospective Register of Systematic Reviews.
QUOROM: Quality of Reporting of Meta-analyses standards.
SIGN: Scottish Intercollegiate Guidelines Network.
SR: Revisiones sistem\'aticas.
RAR: Reducci\'on absoluta del riesgo.
RoB: Riesgo de sesgo.
SJR: SCImago Journal & Country Rank.
UK DCTN: UK Dermatology Clinical Trials Network.

Editor and reviewers questions and investigators responses.

April 18, 2016

Prof. Alex Anstey
Editor
British Journal of Dermatology

Dear Professor,
Thank you for agreeing to consider reviewing manuscript BJD-2016-0277 entitled "Short-term efficacy and safety of new biologic agents targeting IL-23/Th17 pathway for moderate to severe plaque psoriasis: a systematic review and network meta-analysis" for British Journal of Dermatology.
We appreciate the time and efforts in reviewing this manuscript. We carefully considered your comments as well as those offered by the three reviewers, paying closer attention to improve its clarity and flow of ideas. Your valuable feedback information and constructive suggestions will indubitably improve the quality of the manuscript. We have addressed all issues indicated in the review report, and believed that the revised version can meet the journal publication requirements.
Here we enclose the responses to the reviewers.

Yours sincerely,

Juan Ruano
Department of Dermatology
IMIBIC/University of Cordoba
Reina Sofia University Hospital
Córdoba, Spain
P: +34-957-011-214
C: +34-606-887-057
F: +34-957-011-214

Referees' comments:
Reviewer 1:
In this article the authors present a network meta-analysis regarding the short-term efficacy and safety of biologics in psoriasis, including data from secukinumab.
Network meta-analysis has advantages over conventional pairwise meta-analysis, as the technique borrows strength from indirect evidence to gain certainty about all treatment comparisons and allows for estimation of comparative effects that have not been investigated head to head in randomized clinical trials. However, it has some limitations, particularly because it is assumed that all the compared studies are performed under the same standards of populations, designs and outcomes.

The value of short-term efficacy of biologics in psoriasis, despite still interesting, has proven of limited value in decision making from the perspective of clinicians and payers. In example, infliximab is almost invariably the ‘‘most effective’’ drug in the short term but is only scarcely used by dermatologist, as the loss of efficacy with time is increased compared with other drugs.
The article is well written and developed. There are, however, some points to consider.
We greatly appreciate the reviewer’s efforts to carefully review the paper and the valuable suggestions offered. With respect to the concerns raised by the reviewer, we made the following revisions accordingly:

Q1. One may wonder why ixekizumab has been excluded considering that the clinical development has been already completed. Despite it still not strictly licensed for psoriasis, the chances of being approved shortly are very high. The advent of ixekizumab in the next months will probably justify a new meta-analysis with practically the same data very shortly, increasing the number of papers with duplicated content.

R1. We agree with the reviewer that it would be more informative to include data of other biologics which blockade IL-23p19 (guselkumab, tildrakizumab), IL-17A (ixekizumab) or IL-17RA (brodalumab). Our decision was based on two reasons. First, only approved drugs provide estimates of comparative effectiveness that will be potentially useful to current decision-makers, and none of these agents licensed for treatment of psoriasis -except for ixekizumab, approved by the FDA by March 2016, but not yet by the EMA-. Secondly, adressing the 'PROSPERO International prospective register of systematic reviews' rules implies to submit the study protocol to the database at least 6 months before the anticipated analysis completion date -and after that preparing such as extensive paper. This situation entails the risk of that any drug which will be approved beyond PROSPERO final registration -in our case, ixekinumab- may not finally be included in the analysis. Even being aware of such limitation, we consider that submitting in advance all systematic review protocol details to PROSPERO database will undoubtedly help to reduce unplanned duplication and increase transparency, helping safeguard against selective reporting. With this concern, we have now added a paragraph in the Discusion section of the paper (page 16 , lines 12-21).

Q2. One could wonder why to choose ‘‘infectious AE’’ or ‘‘any AE’’ as a marker of safety profile. Infectious AE is by far the most common AE registered in CT. However, in most cases it consist on common colds, usually not related with the drug. Thus, minor differences in the criteria of registration could bias this data. The real weight of ‘‘any adverse event’’ and ‘‘any infection adverse event’’ in the whole safety profile is not clear. From the perspective of decision making, the choice of ‘‘severe adverse event’’ could be more adequate.

R2. We agree with the reviewer that the choice of ‘‘severe adverse event’’ could be more adequate from the perspective of decision making. However, the number of SAEs comunicated in the RCTs of approved drugs is limited -otherwise, a high number of SAEs would impede

its approval by agencies. These kind of AE (SAEs) are mainly found in Phase IV and post-comercial observacional studies (i.e. PsoNet, Biobadaderm). The reason to select ‘‘at least one adverse event (AE)’’ and ‘‘at least one infectious AE’’ safety outcomes to perform the efficacy vs safety analysis was based on the statistically significant differences found among agents for these variables. On the contrary, as no differences were found for 'at least one serious EA' or 'withdrawal owing to AE', it was not meaninful to include them in the analysis.

Q3. The authors detected some inconsistences that could be due to different criteria between the compared studies. It should be detailed in what sense they could condition the results As the authors affirmed, in network meta-analysis the homogeneity of studied populations is particularly crucial. However, there have been important differences in the population studied, from a clear dominance of US American and Canadian population in the first CT, evolving to the recruitment of European, latin American or Asian patients in the later. Average weight, a main conditioning factor in biologic response, is known to be strongly different in these different populations

R3. Thanks for the suggestion. In a systematic review, a decision about whether to pool the results of studies in meta-analysis needs to consider whether there are clinical or methodological differences between studies that might affect the results. We have performed a new analysis in R looking for potential differences in 'average weight' by treatment arms. No significant differences were found (see results bellow). In case of any statistical heterogeneity (I2) was observed, it was always within reasonable limits. Pooled data using general linear model showed little or no variability (Ryan R; Cochrane Consumers and Communication Review Group. ‘Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Review Group reviews: planning the analysis at protocol stage. http://cccrg.cochrane.org, February 2014 (accessed DATE: 02nd February 2016).

```
R version 3.2.2 (2015-08-14)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.9.5 (Mavericks)
attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base
other attached packages:
[1] lme4_1.1-10   Matrix_1.2-2  ggplot2_2.0.0

Call:
glm(formula = AverageWeight_kg ~ group, data = DB_NMA)
Deviance Residuals:
   Min      1Q  Median      3Q     Max
-7.857  -1.417   1.143   2.450   5.143

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)                       8.950e+01  1.978e+00  45.245   <2e-16 ***
groupEtanercept 25mg BIW          2.500e+00  4.423e+00   0.565    0.579
groupEtanercept 50mg BIW          3.571e-01  2.480e+00   0.144    0.887
groupInfliximab 5 mg.kg-1         9.959e-16  3.426e+00   0.000    1.000
groupSecukinumab 300 mg Q4W      -9.000e-01  2.654e+00  -0.339    0.738
groupUstekinumab 45 mg Q12W       1.833e+00  3.022e+00   0.607    0.551
groupUstekinumab 45mg/90mg Q12W -1.000e+00  3.426e+00  -0.292    0.774
groupUstekinumab 90 mg Q12W       2.833e+00  3.022e+00   0.938    0.360
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 15.65213)

    Null deviance: 338.96  on 26  degrees of freedom
Residual deviance: 297.39  on 19  degrees of freedom
  (7 observations deleted due to missingness)
AIC: 159.4
```

Q4. If I have understood, the ‘‘articular’’ case is considered as a separate item. However, considering the heterogeneity of the articular dominions of psoriasis, that has not been, differently to those seen in skin lesions, adequately evaluated, could bias the results.

R4. We agree with the reviewer. We have performed a new analysis in R to assess the presence of between-study variation in 'percentaje of PSA' by treatment arms. No significant differences were found (see results bellow).

```
R version 3.2.2 (2015-08-14)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.9.5 (Mavericks)
attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base
other attached packages:
[1] lme4_1.1-10   Matrix_1.2-2  ggplot2_2.0.0

Call:
glm(formula = PsA_percentage ~ group, data = DB_NMA)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-58.000  -20.167   -0.667   18.750  159.000

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                     69.000     24.136   2.859    0.010 *
```

```
groupEtanercept 25mg BIW         -33.500     41.804  -0.801    0.433
groupEtanercept 50mg BIW          -2.167     31.159  -0.070    0.945
groupInfliximab 5 mg.kg-1          1.000     36.868   0.027    0.979
groupSecukinumab 300 mg Q4W      -25.500     34.133  -0.747    0.464
groupUstekinumab 45 mg Q12W       11.333     36.868   0.307    0.762
groupUstekinumab 45mg/90mg Q12W  -10.500     41.804  -0.251    0.804
groupUstekinumab 90 mg Q12W       25.667     36.868   0.696    0.495
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 2330.114)

    Null deviance: 51438  on 26  degrees of freedom
Residual deviance: 44272  on 19  degrees of freedom
  (6 observations deleted due to missingness)
AIC: 294.48
```

Q5. The authors affirmed that secukinumab and infliximab were the most effective agents. However, it is not clear what is considered ‘‘efficacy’’; was it PASI 75, 90, an average of both?

R5. Thanks for your valuable comment. We state that secukinumab and infliximab were the most effective agents based on Surface Under the Cumulative RAnking probabilities (SUCRA) for PASI 75 (95.9\% for infliximab 5 mg.kg-1 Q8W; 90.0\% for secukinumab 300mg Q8W) and PASI 90 (86.7\% for infliximab 5 mg.kg-1 Q8W; 91\% for secukinumab 300mg Q8W). These treatments were followed by ustekinumab 90 mg Q12W (SUCRA for PASI 75: 73.6\%; SUCRA for PASI 90: 71.7\%). These data are provided in Table 4.

Reviewer 2

Q1. The analysis is both too complex with massive amounts of data presented in supplementary files that, all told, are not very useful to a clinical decision making process.

R1. Thanks for the comments. We admit that the analysis is too complex with massive amounts of data presented in supplementary files. But we consider that performing an extensive and systematic literature search is one of the strengths of any systematic review. And as results, a great amount of data is obtained which must be filtered and analysed. Following well stablished methodology (PROSPERO database, PRISMA extension for NMA, Cochrane rule, GRADE approach) and providing raw data, methodology, results and the software or script of code used for analysis will assure transparency a reproducibility of our results. Clinical decision making is a complex process that implies managing many sources of information. For that reason and in an attempt to facilitate the comprehension of our results several summarizing graphs (eFigures 1-3) and tables (table 4, eTable 1) have been included in the paper.

Q2. Also the analysis is not balanced to include all agents in IL-17 class and emerging IL-23 antagonists (admittedly with much less data available).

R2. We agree with the reviewer that it would be more informative to include data of other biologics. This issue has been addressed above (Reviewer 1, Q2). Briefly, our decision was based on two reasons. First, only approved drugs provide estimates of comparative effectiveness that will be potentially useful to current decision-makers, and none of these agents licensed for treatment of psoriasis -except for ixekizumab, approved by the FDA by March 2016, but not yet by the EMA-. Secondly, adressing the 'PROSPERO International prospective register of systematic reviews' rules implies to submit the study protocol to the database at least 6 months before the anticipated analysis completion date -and after that preparing such an extensive paper. This situation implies the risk of that any drug approved beyond PROSPERO final registration -in our case, ixekizumab- may not be finally included in the analysis. Even being aware of such limitation, we consider that submitting in advance all systematic review protocol details to PROSPERO database will undoubtedly help to reduce unplanned duplication and increase transparency, helping safeguard against selective reporting. With this concern, we have now added a paragraph in the Discusion section of the paper (page 16, lines 12-21).

Q3. The risk of a a small increase in infection risk with infliximab (hardly used now) and secukinumab is not easily framed or understood against potential decreased risk for systemic toxicity and other effects of classic former drugs. Furthermore, a thoughtful consideration of risk/benefit must go to possible benefits of treating psoriasis (with any agent) against serious co-morbidity risks and shortened lifespan that may be impacted by use of good and well tolerated treatments. This piece stands to serve as a "scare tactic" to move therapeutic targeting away from emerging treatments that are emerging as the most effective.

R3. We appreciate the reviewer’s insightful comments and helpful suggestions. We agree that a thoughtful consideration of risk/benefit balance must be taken into account when treating any patient with psoriasis. To do this it is necessary to dispose previously of sufficient and unbiased information for every drug that will be considered. Reporting systematic reviews comparing multiple treatments help us to summarise the available empirical evidence regarding the efficacy and safety of all therapeutic alternatives. Obviously, even of its high quality, evidence obtained from RCT by this approach is not the only source of information to consider in any decision-making process by the patients, physicians, regulators, health technology assessors, and third-party payers. Phase IV clinical trials and experience of use are two useful sources of information (i.e. loss of efficacy over time, lack of efficacy in some populations, new adverse effects not seen in RCTs, etc). Thus, the fact that there are differences in the amount, nature and quality of evidence between old -such as infliximab or adalimumab and etanercept- and new drugs seems obvious. But for the purpose of reporting the results of our systematic review, only evidence obtained from RCTs data using a well known and recognized scientific methodology will assure the reproducibility of unbiased results.

Q4. Yes there are some risks of infections, but the range is narrow (candidasis as a specific signal with secukinumab) and generally treatable.

R4. We agree with your assessment, in the majority of cases, infectious adverse events were moderate or mild. Indeed, severe cases of infection probably were accounted as SAEs and thus being difficult to identify them in most of the RCTs. In our study, we found differences in quality of evidence between eficacy and safety outcomes across pooled RCTs (eFigure 3). In most cases, pooled RCTs tend to score lower for safety as compared with eficacy outcomes. This highlights the lack of quality of adverse event-related data published, which in contrast with the quality of efficacy-related information of these drugs. Even with that limitation, the odds ratio for infections was significantly higher for certain drugs (infliximab 5mg.kg-1 Q8W, adalimumab 40mg Q2W, etanercept 50mg B1W, secukinumab 300mg Q4W) in comparison with placebo. Not taking into account these results would lead to an underestimation -even more- of safety outcomes related with these drugs, a primary objective of any treatment even in the setting of a RCT. Nevertheless, we consider that it would be useful to clarify better this issue. For this reason, we have included in the Discussion section a paragraph making mention on this regard (on page 13, lines 25-26).

Reviewer: 3
This network meta-analysis assesses the short-term effectiveness and safety of new biologics for moderate to severe plaque psoriasis. The team have used a robust methodology to address this question, and the results are reported in a clear format. I have a few comments that could be addressed to improve the manuscript.

R0: Your encouraging comments are greatly appreciated. We greatly appreciate the efforts to carefully review the paper and the valuable suggestions offered. Threfore, we have made the following revisions accordingly.

Q1. It is not clear why 2000 was chosen as the start date for the search strategy - could this information please be added for clarity?

R1. A review of the literature conducted in previous studies (Reich et al. 2012) didn't find any RCT that was performed to assess eficacy and safety of any of the agents considered in our review prior to that date.

Q2. It is not clear who extracted the data from the eligible studies - please clarify whether this was done independently by two reviewers.

R2. As suggested, the statement ‘‘Data were extracted into the open source postgreSQL database by one author (FG) and assessed for accuracy by a second author (JR)’’ has been added to the manuscript (Supplementary material: 1.- Supplementary material and methods).

Q3. About 25\% of the Risk of Bias is categorised as 'unclear', given that the review only included recent studies, it is not clear why the authors of the papers were not contacted for further information.

R3: We agree with the editor. However, some studies suggests that while including previously unpublished data can result in clinically important changes in estimates, the magnitude and direction of impact may not be readily predictable. Vale et al. who compared risk of bias (RoB) in trials with information obtained during IPD analyses found all the unclears were actually low risk of bias. Supplementary information reduced the proportion of unclear assessments for all individual domains, consequently increasing the number of trials assessed as low risk of bias, and therefore available for inclusion in meta-analyses (Vale, et al. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. BMJ 2013;346:f1798). In our case, that we did not to contact with the authors theRoB probably would be of overstimated. For this reason, we have now included in the Discussion section a paragraph making mention on this regard (page 15, lines 16-21). Thank you for point it out.

Q4. It is not clear what the 75 and 90 relate to of the PASI - could this information please be added so it is clearer to the reader that these are recognised cut offs for this scale.

R4. As suggested, the statement ‘‘PASI 75 and PASI 90 represent a 75\% and 90\% or more reduction, respectively, in the PASI score with respect to baseline’’ has been included in the updated version of the manuscript (Materials and Methods section; page 8, lines 1-2). We agree that a more accurate description will make clearer to any reader to understand the significance of PASI outcomes.

Q5. Were any studies excluded due to the outcomes only being reported on a continuous scale or where other cut offs were used, e.g. PASI 50?.

R5. Thanks for the insightful comment suggestion. No study was excluded for such reasons as reflected in the PRISMA extension for NMA 2015 checklist (Supplementary Material, eTable 3) and in the register file submited to PROSPERO prospective repository (Supplementary Material, eTable 4) previously to literature searching and the extraction of data.

Q6. Please add the reasons for why the 55 full text articles were excluded to eFigure 1.

R6. We much appreciate the Editor's careful review. Reasons for 55 full text article exclusion in the sample search strategie now appear in a new version of eFigure 1 (Supplementary Material). These reason were mainly:
55 full-text articles excluded, with reasons
Subanalysis of RCTs (n=14)
Cost-efectivity analysis (n=1)
RCTs in teenagers/children population (n=2)
RCTs comparing treatment schemes (n=16)
RCTs assessing long-term efficacy (n=5)
Phase II RCTs with other molecules (n=1)
CTs in other populations (n=9)
Meta-analysis (n=4)
Excluded for other reasons (n=3)

Q7. Due to the outcomes in the two groups being common, the ORs will overestimate the apparent treatment effects. Therefore there is the strong potential for the results to be misinterpreted. Why were ORs decided as the most appropriate measure of effect to use to this meta-analysis?

R7. The non-equivalence of the OR and RR does not indicate that either is wrong: both are entirely valid ways of describing an intervention effect. Among effect measures for dichotomous data, no single measure is uniformly best, so the choice inevitably involves a compromise. The most important mathematical criterion to select OR is the availability of a reliable variance of the OR that does not rely on which of the two outcome states is coded as the event and is the only statistic which is unbounded. On average there is little difference between the odds ratio (OR) and risk ratio (RR) in terms of consistency (Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Smith GD, Altman DG, eds. Systematic reviews in healthcare: meta-analysis in context. London: BMJ Publishing Group,2001:285-312). Problems may arise, however, if the OR is misinterpreted as a RR. For interventions that increase the chances of events, the odds ratio will be larger than the risk ratio, so the misinterpretation will tend to overestimate the intervention effect, especially when events are common.

Q8. Could the authors please clarify whether they extracted outcomes based on intention to treat.

R8. We extracted outcomes based on intention to treat when authors described the analysis as ITT and with regard to this requirement they included data for all randomized participants in the tables. Per your advice, we now have included a sentence in the method section to clarify it (page 7, line 21).

Q9. The discussion section does not adequately describe the limitations relating to the...
Q9.1. [...]small number of head-to-head comparisons,

Q9.2. [...]the high level of inconsistency for particular comparisons,
Q9.3. [...]nor the impacts of strong evidence of publication bias and imprecision in the majority of active vs placebo comparisons.

R9.1-3. As suggested, we have clarified all these limitations in the new Discusion section.

Q10. Tables on page 47 onwards: It does not seem appropriate to say that there is 'no evidence' of publication bias, when there is only one or two studies assessing a particular comparison.

R10. That is a good question that is frequentely asked. Publication bias generally when there is only one study found is less of a concern when the search has been conducted well, in particular when a systematic search was actually done to eliminate the risk of not identifying studies. As we performed a systematic review, publication bias in those cases should therefore not be assumed by default. That situation does not necessarily downgrade quality of the evidence due to this issue (probably for instance due to risk of bias, imprecision or other factors).

Associate Editor Comments for Authors:
The review authors have have worked really hard on this review. However, in addition to the comments as discussed by the peer reviewers there are several additional comments to be made.

R0: Thank you very much for your kind words about our paper. We appreciate the Associate Editor's insightful comments. In the following sections, you will find our responses to each of your points and helpful suggestions. We are grateful for the time and energy you expended on our behalf.

Q1. The manuscript text is too long and and should not exceed 3000 words of body text.

R1. We have moved par of the Method and Results Section to the Supplementary File in order to meet the journal publication requirement related with total word count of 3.066.

Q2. What this study add ...What we would like to read here is ‘‘What do we now know as a result of this systematic review and network meta-analysis that we did not know before?" At this moment the text is not very informative.

R2. As Editor has suggested, we now provide new bullet statements to highlight the results of our review. Thank you for point it out.

Q3. Abstract: I think abbreviations in abstract Q8W etc needs explanation.

R3. Thank you for this suggestion. In the revision, we deleted abbreviations such as Q8W, Q4W or Q12W because they only appears once in the abstract. Instead of them, the their full expression (every eight weeks, every four weeks, and every twelve weeks) have been written.

Q4. The review authors stated that the systematic review was conducted in accordance with the PRISMA extension for NMA 2015, but no results are reported regarding efficacy and safety in the abstract (ORs and CI) (Item 2 on the checklist as published in Hutton).

R4. We are sorry for the confusion. We have now included OR and 95\%CI data for efficacy and safety in the abstract and in the PRISMA extension for NMA 2015 checklist.

Q5. Abstract: Cochrane compliant rules...quality of evidence follows GRADE approach (see also under Methods: quality of evidence assessment).

R5. As requested we have changed the wording to ‘‘quality of evidence follows GRADE approach’’ in Abstract (on page 4, line 6) and in 'quality of evidence assessment' subheading of the new Supplementary Materials and Methods in the revised version.

Q6. Introduction: 2nd paragraph it is strength of recommendation and quality of evidence, not strength of evidence.

R6. Thank you for this suggestion. In the revision, we have changed ‘‘Meta-analyses are conducted to assess the strength of evidence available for a disease and multiple treatment alternatives[...]’’ by ‘‘Meta-analyses are conducted to assess the strength of recommendation and quality of evidence available for a disease and multiple treatment alternatives[...]’’ on Introduction section (page 5, lines 15-16).

Q7. There is no mentioning that only studies in English literature would be included (language bias).

R7. You raise an important question regarding potential language bias. Reviews are often exclusively based on trial published in English, and ours is also. The proportion of controlled trials with statistically significant results seems to be higher among reports published in English. Language bias could thus be introduced in those meta-analysis exclusively based on English-language reports. Nevertheless, large bibliographic databases, such as MEDLINE and EMBASE which were used by our team, do include a small number of non-English language journals. There are two mentions about this issue in the Materials and Method section (page 7, line 8) and in the PRISMA NMA Checklist document (eTable 3).

Q8. The Discussion would benefit from headings such as strenghts of this review and limitations of this review.

R8. That’s an good suggestion to improve reading comprehension. We have now included ‘‘Strenghts of this review’’ and ‘‘Limitations of this review’’ headings in the 'Discusion' section of the current version of the manuscript.

Q9. The SOF tables: The review authors have downgraded the quality of evidence for imprecision at several places, but then you cannot upgrade the quality of evidence!
GRADE Handbook 5.3
"Note: Consideration of factors reducing quality of evidence must precede consideration of reasons for rating it up. Thus, the 5 factors for rating down quality of evidence (risk of bias, imprecision, inconsistency, indirectness, and publication bias) must be rated prior to the 3 factors for rating it up (large effect, dose-response and effects of residual confounding). The decision to rate up quality of evidence should only be made when serious limitations in any of the 5 areas reducing the quality of evidence are absent.
Note: Although it is theoretically possible to rate up results from randomized control trials, we have yet to find a compelling example of such an instance."

R9. We appreciate your comments. As suggested, SOF tables and scoring heatmap of quality of evidence for each outcome across pooled studies have been updated in relation to imprecision and inconsistency assessment in the new version of the review. Briefly:
For 'PASI 75':
-Quality of evidence 'infliximab 5mg.Kg-1 Q8W vs placebo' was downrated to moderate due to serious limitations in imprecision (strong/very strong association was not considered);
-Quality of evidence 'etanercept 50mg BIW vs placebo' was downrated to moderate due to serious limitations in imprecision (strong/very strong association was not considered);
-Quality of evidence 'etanercept 25mg BIW vs placebo' was downrated to low due to very serious limitations in imprecision (strong/very strong association was not considered);
-Quality of evidence 'adalimumab 40mg Q2W vs placebo' was downrated to moderate due to serious limitations in imprecision (strong/very strong association was not considered);
-Quality of evidence 'secukinumab 300mg Q4W vs placebo' was downrated to moderate due to serious limitations in imprecision (strong/very strong association was not considered);

For 'PASI 90':
-Quality of evidence 'infliximab 5mg.Kg-1 Q8W vs placebo' was downrated to moderate due to serious limitations in imprecision (strong/very strong association was not considered);
-Quality of evidence 'etanercept 50mg BIW vs placebo' was downrated to low due to serious limitations in inconsistency and imprecision (strong/very strong association was not considered);
-Quality of evidence 'etanercept 25mg BIW vs placebo' was downrated to low due to very serious limitations in imprecision (strong/very strong association was not considered);
-Quality of evidence 'adalimumab 40mg Q2W vs placebo' was downrated to very low due to serious limitations in imprecision (strong/very strong association was not considered);
-Quality of evidence 'ustekinumab 45mg Q12W vs placebo' was downrated to moderate due to serious limitations in inconsistency (strong/very strong association was not considered);

Quality of evidence ' ustekinumab 90mg Q12W vs placebo' was downrated to moderate due to serious limitations in inconsistency (strong/very strong association was not considered);
-Quality of evidence 'secukinumab 300mg Q4W vs placebo' was downrated to moderate due to serious limitations in imprecision (strong/very strong association was not considered);

For 'IGA/PGA/sPGA=0/1':
-Quality of evidence 'infliximab 5mg.Kg-1 Q8W vs placebo' was downrated to low due to serious limitations in inconsistency and imprecision (strong/very strong association was not considered);
-Quality of evidence 'etanercept 50mg BIW vs placebo' was downrated to moderate due to serious limitations in inconsistency (strong/very strong association was not considered);
-Quality of evidence 'adalimumab 40mg Q2W vs placebo' was downrated to very low due to serious limitations in inconsistency and imprecision (strong/very strong association was not considered);
-Quality of evidence 'secukinumab 300mg Q4W vs placebo' was downrated to very low due to serious limitations in inconsistency and very serious limitations in imprecision (strong/very strong association was not considered);

For DLQI=0/1:
-Quality of evidence 'adalimumab 40mg Q2W vs placebo' was downrated to very moderate due to serious limitations in inconsistency and imprecision (strong/very strong association was not considered);

Q10. For other comparisons the review authors have NOT downgraded for imprecision while the optimal information size is not met and there is low occurrence of events eg secukinumab 300 mg Q4W vs ustekinumab 45/90 mg Q12W at least one infectious adverse event See GRADE guideline 6 and sometimes you have to downgrade twice for imprecision (very few events).

R10. Author is grateful to the editor for these valuable comments. Following your suggestions imprecision assessment has been have been reviewed and updated in al SOF tables and scoring heatmap of quality of evidence for each outcome across pooled studies of the paper. Briefly:

For 'ustekinumab 45mg Q12W vs. placebo', imprecision was downrated to very serious for ‘‘withdrawal due to adverse event’’.
For 'secukinumab 300mg Q4W vs. placebo', imprecision was downrated to very serious for ‘‘withdrawal due to adverse event’’.
For 'etanercept 50mg BIW vs. ustekinumab 45mg Q12W':
- imprecision was uprated to not serious for ‘‘at least one adverse event’’;
- imprecision was downrated to very serious for ‘‘at least one serious adverse event’’;
- imprecision was downrated to very serious for ‘‘withdrawal due to adverse event’’.
For 'etanercept 50mg BIW vs. ustekinumab 90mg Q12W':
- imprecision was downrated to very serious for ‘‘at least one infectious adverse event’’;

- imprecision was downrated to very serious for ‘‘at least one serious adverse event’’;
- imprecision was downrated to very serious for ‘‘withdrawal due to adverse event’’.

For 'etanercept 50mg BIW vs. secukinumab 300mg Q4W':

- imprecision was downrated to very serious for ‘‘at least one serious adverse event’’;
- imprecision was downrated to very serious for ‘‘withdrawal due to adverse event’’.

Q11. I would like to suggest to submit the Checklist as listed in the paper of Hutton and add at which page the items are addressed.

R11. Thanks for the comments. We have included as suggested the PRISMA extension for NMA checklist on the Supplementary Material as eTable 3. Pages in the manuscript on which items of the checklist were addressed have been included too.

Finally, we wanted to thank you again for the careful reviewing and the positive and constructing comments, which we deeply thank.

January 12, 2017

Prof. Alex Anstey
Editor
British Journal of Dermatology

Dear Professor,
Thank you for agreeing to consider reviewing manuscript BJD-2016-2340 entitled "Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest, and bibliometric indices as predictors of methodological quality" for British Journal of Dermatology.
We appreciate the time and efforts in reviewing this manuscript. We carefully considered your comments as well as those offered by the two reviewers, paying closer attention to improve its clarity and flow of ideas. Your valuable feedback information and constructive suggestions will indubitably improve the quality of the manuscript. We have addressed all issues indicated in the review report, and believed that the revised version can meet the journal publication requirements.
Here we enclose the responses to the reviewers.

Yours sincerely,

Juan Ruano
Department of Dermatology
IMIBIC/University of Cordoba
Reina Sofia University Hospital
Córdoba, Spain
P: +34-957-011-214
C: +34-606-887-057
F: +34-957-011-214

Referees' comments:
Reviewer 1:
General comments:
There has certainly been an explosion of systematic reviews in the field of dermatology and many for psoriasis. Understanding the factors that might indicate a higher quality systematic review is useful. The findings that high bibliographic scores for certain authors may be related to poorer methodological quality is interesting and suggests that their names are being hired to influence readers. The analysis and data presentation is thoughtful and so is the discussion.
We greatly appreciate the reviewer’s efforts to carefully review the paper and the valuable suggestions offered. With respect to the concerns raised by the reviewer, we made the following revisions accordingly:

Major
Q1. It is not clear at all in the introduction why you have chosen to only look at psoriasis.

R1. The study was conducted by a research group that it is focused on immune-mediated inflammatory skin diseases. Moderate-to-severe forms of psoriasis are associated to significant comorbidity, impaired quality of life, and high direct and indirect costs, thus therapeutic decision-making process may imply both clinical and economic factors. In addition given the large number of systematic reviews we think that to make the work feasible it is necessary to limit the subject matter. We have recently performed a systematic review and network meta-analysis as a part of a research project on psoriasis that has received competitive public academic funding. We have included a brief paragraph in the Introduction section of the manuscript as suggested (on page 6, lines 2-4).

Q2. Why did you hypothesise that industry studies are more likely to have methodological bias than other studies? - industry often spends a lot of money making sure that their methods and reporting are good. It is framing bias (the way the data is described in a positive light and some data such as drug harms are concealed or reduced) that they are famous for. Pity you did not look at framing bias in such reviews.

R2. It is known that outcome reports of RCTs and MAs funded by the pharmaceutical industry are more likely to favor the sponsor’s product compared to studies with other sources of funding (Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration.J Clin Epidemiol. 2009 Oct;62(10):e1-34). AMSTAR measures SR quality by fundamentally reviewing methodological aspects of them. For example, one of AMSTAR's items is 'publication bias', and in this sense it is well known that trials promoted by pharmaceutical companies with negative results tend not to be published (Song F, Lee H, Yoon KL, Publication bias: What is it? How do we measure it? How do we avoid it? Open Access Journal of Clinical Trials 2013: 5 71-81). If we assume that the tendency to have positive results in the SRs/MAs financed by the industry is artificial a plausible explanation would be that the methods for achieving biased results may be altered. A tool such as AMSTAR should find differences in SRs/MAs results based on the source of funding. Thanks for the suggestion to look at framing bias in future post-hoc analysis.

Q3. What was the rationale for dividing AMSTAR scores in low, moderate and high? Presumably there is a literature using anchor based methods that you can cite to support such cut-offs?

R3. One of the criticized aspects of AMSTAR is that there is no provided guidance on how to translate the total score into categorical ratings (Burda BU, Holmer HK, Norris SL. Limitations of A Measurement Tool to Assess Systematic Reviews (AMSTAR) and suggestions for improvement. Syst. Rev. 2016;5:1-10.). However, most of studies establish quality levels using similar cutoff points for low (0-3), moderate (4-7), and high methodological

quality (8-11) respectively. We have been stricter in distinguishing between high quality SRs from moderate and low SRs.

Q4. A real pity that you restricted this review to English language only studies, which as you are aware is marker of poor quality for SRs because of a documented language bias.

R4. You raise an important question regarding potential language bias. We limited to English language reviews because of time limitations for project completion. That reason was explained at the Method and Discussion sections and was highlighted at protocol vs overview comparison. Reviews are often exclusively based on trial published in English, and ours is also. The proportion of controlled trials with statistically significant results seems to be higher among reports published in English. Language bias could thus be introduced in those meta-analysis exclusively based on English-language reports. Nevertheless, large bibliographic databases, such as MEDLINE and EMBASE which were used by our team, do include a small number of non-English language journals. Despite everything, we have a large number of studies that we consider representative (n=220), and in any case, the ability of the model to identify the studies of low quality was high [Results of k-fold cross-validation demonstrated that our model performed better to predict low vs. moderate or high quality reviews (κ= 0.319; sensitivity: 0.83, specificity: 0.76; positive predictive value: 0.50; negative predictive value: 0.94].

Q5. Utility of the review: Although such aggregate conclusions will not serve as a substitute for assessing review quality using a suitable instrument like AMSTAR or ROBIS on a case by case basis. Please include something on the challenge of aggregate data in your discussion so that you encourage readers to look at each study in its own merit.

R5. Thanks for the comments. We admit that the analysis is too complex with massive amounts of data presented in supplementary files. But we consider that performing an extensive and systematic literature search is one of the strengths of any systematic review. And as results, a great amount of data is obtained which must be filtered and analysed. Following well stablished methodology and providing raw data, methodology, and the scripts of code used for analysis will assure transparency a reproducibility of our results. We agree that is probably most useful development tools such as a decision tree or a new version of AMSTAR with questions weighted on the basis of our results. Until then, it would be risky to advise readers of using any combination of the identified factors without previously knowing the power of these tools and carring out internal and external validation assessments. In addition, it would be desirable testing if our results replicate when reviews in other research topics (i.e., areas less influenced by the pharmaceutical industry, such as rare diseases). Finally, comparing the capacity of ROBIS vs. AMSTAR to classify such revisions could be an interesting future research project (PROSPERO ID: CRD42016053181). With this concern, we have now added a paragraph in the Discusion section of the paper (on page 19, lines 21-25, and page 20, lines 1-4).

Q6. Even though it sounded as if you were out to expose industry bias, the actual magnitude of that effect (as measured by number of authors with conflict of interest) was quite small (OR 0.9). So the conclusions need to acknowledge this small effect more (although I suspect it would have been a lot greater for measuring framing bias). At the moment (especially in the section what does this study add) it sounds as though you have uncovered major industry bias which at best is small in your study compared to the other factors.

R6. Thanks for your valuable comment, that's an good suggestion. Per your advice, we now have modified the second paragraph of the ''what does this study add?'' section: ''As a summary, when reviews are funded by pharmaceutical academic institutions, authored by few researchers with conflict of interest, or meta-analyses are included in the study, the probability of low methodological quality decreases, reducing the risk of bias of SRs and MAs'' (on page 3, lines 20-23).

Minor
Q7. Given that other studies have compared the quality of Cochrane systematic reviews to non-Cochrane systematic reviews in dermatology, it would be useful to assess the relationship between Cochrane production and study quality. Please consider including Cochrane vs non-Cocgrane as a ''risk factor'' for SR quality in a post hoc analysis.

R7. That's an good suggestion. Cochrane reviews are among the papers with the highest quality in our study and have contributed to define the profiles of results obtained. We will take this suggestion into account for further analysis we are planning -SR quality assessments with others tools different from AMSTAR such as ROBIS tool.

Q8. In the summary, the sentence ''and number of authors with a conflict of interest (OR, 0.9; 95\% CI, 0.824-0.985) significantly predicted a higher quality'' is an odd one as number of authors predicts lower quality. I think you need to disaggregate the +ve from the -ve markers.

R8. As requested we have changed the expression ''and number of authors with a conflict of interest (OR, 0.9; 95\% CI, 0.824-0.985) significantly predicted a higher quality'' by ''[...] and article page count (OR, 1.08; 95\% CI, 1.02-1.15) significantly predicted a higher quality; high number of authors with a conflict of interest (OR, 0.9; 95\% CI, 0.824-0.985) was significantly associated with a lower quality.'' in the revised version (on page 5, lines 18-21). That will undoubtedly improve reading comprehension of the paragraph.

Q9. Page 8 The sentence ''Metaanalyses (MAs) enable the quantitative synthesis of SRs'' is not quite right - I think the authors means that ''MAs enable the quantitative syntheses of randomised controlled trial where appropriate''

R9. We appreciate your comment. As suggested, the expression ‘‘Metaanalyses (MAs) enable the quantitative synthesis of SRs’’ has been changed by ‘‘Meta-analyses (MAs) enable the quantitative syntheses of randomised controlled trial where appropriate’’. This modification is highlighted in the new version of the review (on page 6, lines 11-12).

Q10. Methods: how many people screened possible titles to decide if they were genuine SRs and how did you define an SR? (sometimes it is difficult eg if only one database was searched)

R10. All the studies were selected by two reviewers. Some papers generated discussion and were included by agreement. Since AMSTAR evaluates the quality of the SRs and one of the items is about the analysis performed on them, we only select the papers, which in our opinion, either in the methodology section or in the supplementary material provided sufficient information that ensured that a systematic review was carried out.

Q11. I suggest that you also rate your own review with AMSTAR. If you don’t do it now, I am sure somebody will do it for you in the correspondence when the article is published!

R11. Thank you for this suggestion. We have calculated the AMSTAR score for the article ‘‘Gómez-García F, Epstein D, Isla-Tejera B, et al. Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis. Br J Dermatol. 2016 Jun 13.

AMSTAR item
Answer
Score
Q1: Was an 'a priori' design provided?
Yes:1
Q2: Was there duplicate study selection and data extraction?
Yes:1
Q3: Was a comprehensive literature search performed?
Yes:1
Q4: Was the status of publication (i.e. grey literature) used as an inclusion criterion?
Yes: 0
Q5: Was a list of studies (included and excluded) provided?
No (lack of excluded list) :0
Q6: Were the characteristics of the included studies provided?
Yes:1
Q7: Was the scientific quality of the included studies assessed and documented?
Yes:1
Q8: Was the scientific quality of the included studies used appropriately in formulating conclusions?
NA:0

Q9: Were the methods used to combine the findings of studies appropriate?
Yes:1
Q10: Was the likelihood of publication bias assessed?
Yes:1
Q11: Was the conflict of interest included?
Yes:1
Total:9

Q12. Did you by any chance see if the same bunch of conflicted authors names popped up time and time again for the lower quality industry studies (rent a key opinion leader syndrome). I can tell you who they might be, but it would be interesting to find out.

R12. Thanks for your valuable comment. You raise an important question regarding potential source of bias and ethical conflict in research. We have started a new analysis of author-paper affiliation network architecture based on this data. Differences of collaboration patterns between low, moderate and high methodological quality of reviews will be explored at author, institution, and country level and any potential deviation of the balance between of author-derived factors such as productivity (number of paper by author) or scientific quality (author's h-index) will be assessed.

Q13. I think you could make more of the fact that only 17\% of the studies were quality and that not all SRs are the same.

R13. Thanks for your valuable comment. We have included a brief paragraph in the final conclusion highlighting this important lack of quality among the reviews in psoriasis (on page 20, lines 6-8).

Reviewer 2
COMMENTS TO AUTHORS
Q1. You included some papers as metaanalyses which were editorials, commentaries or methodological papers and not appropriate for scoring.
R1. The reviews were selected by two authors. Some generated discussion and were included by agreement. Since AMSTAR evaluates the quality of the systematic reviews and one of the items is about the analysis performed on them, we only select the SRs, which in our opinion, either in the text or in the supplementary material provided sufficient information on the methodology of the systematic review carried out. However most of the reviews included are SRs. A list of included and excluded papers is supplied with the Supplementarial Material.

Q2. Some of the most interesting findings are in supplemental Tables 6 and 7, especially what seems to be an association between high numbers of publications from an institution and low quality. The country and institutional concentration of such high number poor

scores is interesting. Perhaps, comparing institutions with poor scores number of reviews with their number of original psoriasis research publications in journals with impact factors >3 would be of interest. Looking at ownership of journals that publish poor quality reviews ie proportion that are essentially junk mail and published by for profit company without relevant association with an established professional organization would be of interest.

R2. Thanks for the comments. We have started a new analysis of author-paper affiliation network architecture considering thoses suggestions and based on this first article. Differences of collaboration patterns between low, moderate and high methodological quality of reviews will be explored at author, institution, and country level. Influence of factors such as productivity (number of paper by author) or scientific quality (author's h-index) will be assessed.

Associate Editor Comments for Authors:
This is a very comprehensive review of reviews. I agree with the peer reviewers that it is a pity that the search was limited to English literature, but that was explained, and the amount of studies found is already huge as well as the wort hat came with that. As the peer reviewers provided already quite extensive comments, I only have little to add:

Q1. General: references need to be in super script behind the full stop, not before.

R1. As Editor has suggested, we now provide references as super script behind the full stop. All changes have been highlighted throughtout the new version of the manuscript. Thank you for point it out.

Q2. Page 7 line 18, I think RCTs is better abbreviation for randomised controlled trials throughout review this is widely accepted.

R2. We appreciate the reviewer’s insightful suggestions. In the new revision, we used RCTs as abbreviation of randomised controlled trials -instead of Cts. All changes have been highlighted in yellow.

Q3. Do the authors don’t feel their conflict of interested affected the results? Could have been discussed as several authors have a conflict of interest with companies that market drugs for psoriasis. How was COI avoided during review process?

R3. You raised an important question regarding potential bias of results due to conflict of interest of some of the authors. We did not planned how to avoid this issue, but there some facts that may be considered and that can argue minimal effect on the results: 1.-Most of authors (67\%) did not have any conflict of interest; those who declared conflict of interest was due to relationship with pharma industry (all three with all pharma companies involved in psoriasis) in the past; 2.-For data extraction and methodological quality

assessment tasks, at least one author without conflict of interest was involved; 3.-This review did not get any funding from any pharmaceutical company; 4.-Transparency has been a main objetive in our study: following well stablished methodology (a comprehensive protocol previously published at PROSPERO, a validated AMSTAR tool) and providing raw data, methodology, results and the script of code used for analysis will ensure transparency a reproducibility of our results.

Finally, we wanted to thank you again for the careful reviewing and the positive and constructing comments, which we deeply thank.

January 30, 2017

Prof. Alex Anstey
Editor
British Journal of Dermatology

Dear Professor,
Thank you for agreeing to consider reviewing a revised version of manuscript BJD-2016-2340.R1 entitled "Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest, and bibliometric indices as predictors of methodological quality" for publication in the British Journal of Dermatology.
We thank the editor and anonymous reviewers for their constructive comments, that have helped us to greatly improve the manuscript. Below, we address all of the reviewers' comments point-by-point, and present our subsequent modifications. All suggested changes to wording and punctuation have been incorporated.
On behalf of all co-authors,

Juan Ruano
Department of Dermatology
IMIBIC/University of Cordoba
Reina Sofia University Hospital
Córdoba, Spain
P: +34-957-011-214
C: +34-606-887-057
F: +34-957-011-214

British Journal of Dermatology
Decision Letter (BJD-2016-2340.R1)
From: bjd@bad.org.uk

To: juanruanoruiz@mac.com
CC:
Subject: Reject with resubmission possible - BJD-2016-2340.R1
Body:
Re: BJD-2016-2340.R1
Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest, and bibliometric indices as predictors of methodological quality

Dear Dr Ruano

Thank you for sending us your paper, which has now been through the Editorial process. I am afraid I cannot offer high enough priority to your paper for us to accept it for publication in the British Journal of Dermatology, certainly not in its current format.

If you do decide to resubmit, the please submit your revised manuscript online at https://mc.manus
Your manuscript will be listed in your Author Centre under 'Manuscripts with Decisions'.
It is very important that you:
1) Underline any areas of the text that you change.
2) Respond to any referee comments (and any additional comments made above) before clicking the manuscript title to begin uploading your revised manuscript file(s).

Yours sincerely,

Professor Alex Anstey
Editor
British Journal of Dermatology

Referees' comments:

Reviewer: 1
COMMENTS TO AUTHORS
I still have the following concerns about the responses to my previous comments:

Q1. The actual changes made in the paper are few, and they contain sloppy errors, which makes me lose confidence in the care in which the main review itself has been done. For example:
-Page 5 line 51 of the pdf ‘‘that only included randomised, randomised controlled trials (RCTs) it can..’’
-Page 21 line 1 ‘‘mehodological quality’’ and ‘‘carring out’’
-Summary line 18 ‘‘5-year impact factor (OR, 95\% CI, 1.02-1.14),’’ - the odds ratio is missing!

Response 1: Thank you for your helpful comments. The errors noted above have been corrected and are underlined in the new version of the manuscript.

Q2. The summary is now clearer - thank you, but the changes made to ‘‘What does this study add’’ section ‘‘As a summary, when reviews are funded by pharmaceutical academic institutions, authored by few researchers with conflict of interest, or meta-analyses are included in the study, the probability of low methodological quality decreases, reducing the risk of bias of SRs and MAs.’’ has really confused me again.
What on earth are ‘‘pharmaceutical’’ academic institutions? -the summary was clear that it was just ‘‘funding by academic institutions’’ that predicted of higher quality. Some of the data in the results suggested that some of those funded by pharma companies were poorer quality, so the mixing up of academic funding and Pharma is not helpful in this key paragraph.
Also in this ‘‘what does this study add’’ section, the direction of the summary statements about which factors increase study quality are not consistent with the summary and the phrase ‘‘probability of reducing low methodological quality’’ is cumbersome - why not just say ‘‘predicts higher quality’’ as in your summary.
Here is a possible rewrite of your addition to the ‘‘What does this study add’’ section:
‘‘We found that higher quality reviews included a meta-analysis, were funded by academic institutions, those with fewer authors, and those that contained a high article influence score. Reviews that contained a high number of authors with conflicts were of lower quality’’
ie keep it simple and consistent with what you said in the summary

Response 2: We wholeheartedly agree with the reviewer’s comment, and have revised this part of the ‘‘What does the study add’’ section of the document as follows (page 3, lines 20-23).
‘‘We found that higher quality reviews included a meta-analysis, were funded by academic institutions, had fewer authors, and had a high article influence score. Reviews that contained a high number of authors with conflicts of interest were of lower quality.’’

Q3. With regards to my previous major comment 3 - please add a mention and reference in the methods section reference to another study where these AMSTAR cut-off scores have been used, and acknowledge in the study limitations section of the discussion that the validity of translating the AMSTAR scores into categories is still unclear, citing the useful Burda et al study that you quoted in your reply.

Response 3: Thank you for the helpful suggestions. We have now added a mention and referenced in the Methods section of the Supplementary Material two different studies where AMSTAR cut-off scores have been previously used. We have also commented in the 'limitations subsection' of the manuscript's Discussion section that the validity of translating the AMSTAR scores into categories is still unclear, and cited the Burda reference (page 17,

lines 20-25; page 18, line 1). All changes are underlined in the new version of the paper and the following has been added to the Discussion section.
‘‘One criticism of AMSTAR is that no guidance has been provided on how to translate the total score into categorical ratings. Various thresholds have been used to define categories for quality, making it difficult to compare assessments across reviews.50 Thus, the validity of translating the AMSTAR scores into three categories (high, moderate, and low methodological quality) is still unclear. Burda et al. recommended adding new items and modifying existing items to assess the quality of the body of evidence and to address subgroup and sensitivity analyses.50 ’’

Q4. In relation to my previous major comment 5, please specifically mention and reference the ROBIS tool in your discussion as it would be very odd not to do so given its prominence and developmental rigour. I should add that I do not have any conflict with the ROBIS tool.

Response 4: We have now specifically mentioned and referenced the ROBIS tool. The following paragraph has been included in the Discussion section (page 18, lines 16-20).
‘‘Comparing the capacity of AMSTAR vs ROBIS to classify such revisions would be an interesting future project, given the prominence of ROBIS as a new rigorous tool. The ROBIS tool was developed to employ accurate methodology across four wide groups of reviews within health-care settings: aetiology. interventions, diagnosis, and prognosis.52 (Whiting P, Savovic J, Higgins JPT, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol 2016;69:225-34)’’.

Q5. In relation to my previous minor comment 7, please include an analaysis of Cochrane vs non-Cochrane review as a predictor of quality. I can understand that many of the other author-specific factors that we have mentioned that could be explored will be the topic of a future study which is fair enough, but looking at Cochrane vs non-Cochrane is such a simple and obvious thing to do that it would be odd not to include it in a paper that explores large aggregate factors of review quality.

Response 5: We have performed a new analysis including Cochrane (n = 6) vs. non-Cochrane (n = 214) reviews as a potential predictor of methodological quality. Figure 3 now includes this new variable and clearly all Cochrane reviews, all of high methodological quality, belong to cluster #6. This cluster is characterized by reviews with a low number of authors and institutions, a low number of authors with a conflict of interest, and reviews funded by academic institutions. However, Cochrane affiliation was not a predictor of a high methodological quality of reviews in our study. The reason may be due to the low number of Cochrane reviews found in our dataset. All Cochrane reviews were of high methodological quality, but not all reviews of high methodological quality were authored by Cochrane members. Many reviews of high quality belonged to cluster #3. These results are presented

in the Discussion section of the revised version of the manuscript as follows (page 18, lines 22-25; page 19, lines 1-6).
‘‘Some studies have found that the Cochrane Collaboration authorship affiliation is a predictor of the methodological quality of SRs and MAs.53-55 Although this subgroup of reviews achieved the highest AMSTAR scores in our study, we did not observed this feature. This may be due to the fact that most of these studies performed a linear regression using the total AMSTAR score as a predicting variable. The scarce number of Cochrane SRs of psoriasis found in our search and the low proportion of total SRs of high methodological quality in our dataset may be two additional factors that can explain this discrepancy. In any case, all Cochrane reviews were found in the cluster of studies with the highest AMSTAR score when PCA was performed.’’

Q6. My previous point 10. Please add a short sentence in the methods that two reviewers screened the searches and provide a clear definition of how they decided the study was a systematic review in the methods of the paper. I notice that reviewer 2 question 1 asked for the same clarity. Defining what a systematic review is when searching is not always easy, and your paper allows others to learn from how you did it.

Response 6: We have now described better how we decided that a study was a systematic review. A paragraph has been included in the Method section of the Supplementary Material as follow (page 2, lines 13-22)).
‘‘In the first stage, abstracts downloaded from the literature searches were screened by two reviewers. Reports of systematic reviews were considered eligible for inclusion if the terms or phrases ‘‘systematic review’’, ‘‘meta-analysis’’ or ‘‘overview’’ were used in the title or abstract, or if the main text provided a clear indication that a systematic review had been carried out. Any study clearly not meeting the eligibility criteria was rejected. In the second stage, full papers were retrieved for the selected candidate study and assessed by two reviewers to identify all SRs and MAs meeting the eligibility criteria. Since AMSTAR evaluates the quality of the SRs and one of the items is about the analysis performed on them, we only select the papers, which in our opinion, either in the methodology section or in the supplementary material provided sufficient information that ensured that a systematic review was carried out.’’

Reviewer: 2
COMMENTS TO AUTHORS
(There are no further comments.)

Associate Editor Comments for Authors:
Q1. I am in agreement with the reviewers that the earlier made comments did not result in the changes that were expected. I would like to recommend the authors to go over the initial comments again as well as the latest comments, and try to address the comments more in full and making the required changes with attention to detail.

Response 1: We totally agree with both reviewer #1 and the Editor. We would like to thank both reviewers for their insightful comments, and for your very and careful review of our paper. After completion of the suggested edits, the revised manuscript has tremendously benefitted from an improvement in the overall clarity of the presentation.

August 14, 2018

A. Knottnerus and P. Tugwell
Editors
Journal of Clinical Epidemiology

Dear Editors:

Thank you for agreeing to consider reviewing a revised version of manuscript JCE_2017_334 entitled "Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool" for publication in the Journal of Clinical Epidemiology.
We thank the editor and anonymous reviewers for their constructive comments that have helped us to greatly improve the manuscript. Below, we address all of the reviewers' comments point-by-point and present our subsequent modifications. All suggested changes to wording and punctuation have been incorporated.
On behalf of all co-authors,

Juan Ruano
Department of Dermatology
IMIBIC/University of Cordoba
Reina Sofia University Hospital
Córdoba, Spain
P: +34-957-011-214
C: +34-606-887-057
F: +34-957-011-214

Ref: JCE_2017_334 Title: Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool Journal: Journal of Clinical Epidemiology

Dear Dr. Ruano,
Thank you for submitting your manuscript to Journal of Clinical Epidemiology. I have received comments from reviewers on your manuscript. Your paper should become acceptable for publication pending suitable minor revision and modification of the article in light of the appended reviewer comments.

I look forward to receiving your revised manuscript as soon as possible.

Kind regards,
Ms Germeraad-Uriot Maastricht Editorial Office Journal of Clinical Epidemiology

---

Reviewer 1

Q1: Abstract: Results: Your statement ‘‘A high risk of bias was detected for most SRs classified as displaying high or moderate methodological quality by AMSTAR’’ seems contradictory to me. In the recent years, most of methodological quality appraisal tools are based on the Risk of Bias (RoB) framework. So, the RoB is an approach to appraise methodological quality. Therefore, your statement seems to be a conflicting one from both a conceptual and an empirical point of view. Conclusions: the aforementioned is applicable to this section.

R1: Although there is a discussion about both concepts in the literature, we have employed the concepts of risk of bias (RoB) and methodological quality included in the Cochrane Handbook1. In this sense, ‘‘The ‘assessment of methodological quality’ suggests an investigation of the extent to which study authors conducted their research to the highest possible standards. This would be different from the RoB or risk of performing a systematic error in conducting the study. That is why although the methodological quality ‘‘[...] has been used extensively in the context of systematic review methods to refer to the critical appraisal of included studies’’, there some reasons commented in the Cochrane Handbook would help to differentiate both concepts:
1) ‘‘[...] the key consideration in a Cochrane review is the extent to which results of included studies should be believed. Assessing risk of bias targets this question squarely.’’;
2) ‘‘[...] a study may be performed to the highest possible standards yet still have an important risk of bias’’;
3) ‘‘Some markers of quality in medical research, such as obtaining ethical approval, performing a sample size calculation, and reporting a study in line with the CONSORT Statement, are unlikely to have direct implications for risk of bias’’;
4) ‘‘An emphasis on risk of bias overcomes ambiguity between the quality of reporting and the quality of the underlying research’’.

Q2: Main text: Line 22: Please review English.

R2: The above mentioned sentence has been eliminated. A new one has been included as follow: ‘‘However, some research groups have found discrepancies in results between reviews designed to response the same research question.’’ (pg 2, ln 20-23).

Q3: Line 32: I am not aware of this ‘‘Multiple quality assessment tools have been developed to assess the methodological quality of reviews’’ and the reference you did not enlighten me. Can you provide more information of these tools and why you didn’t chose some more of them?

R3: Although this statement is included in the document of the ROBIS tool, its basis procedes from the pre-ROBIS Technical Report2, where authors perform a systematic review about tools to assess the quality of this documents of synthesis. They carry out an evaluation of 40 checklists for quality assessment of systematic reviews or meta-analyses, concluding that ‘‘[...] only three had been rigorously developed’’. Of all, they state that ‘‘although there is currently no accepted tool to assess the quality of systematic reviews, our preliminary searches demonstrated that the AMSTAR tool is the most commonly used’’. In addition, AMSTAR instrument has proven to be a valid and reliable tool in the evaluation of the methodological quality of systematic reviews3. Finally, our research group has previously being experienced with the management of this tool4. We thank the reviewer for the suggestion. A new cite (# 10) to the above mentioned technical report has been included in a new version of the manuscript to support our statement (page 8, lines 621-625).

Q4: Line 49. Any study to support this statement? I can suggest you a recent one: Oliveras, I., Losilla, J.-M., Vives, J. (2017).Methodological quality is underrated in systematic reviews and meta-analyses in health psychology.Journal of Clinical Epidemiology,86, 59-70.https://doi.org/10.1016/j.jclinepi.2017.05.002

R4: This assertion derives from conceptual differences between RoB and methodological quality. As we have previously discussed, high standards of methodological quality do not eliminate the RoB.
We have included the suggested cite (# 13) in a new version of the manuscript to support this statement (page 8, lines 633-636).

Q5: Line 81-83. I am not sure of the influence of one on the other. Any study to support this statement?

R5: Thank you very much for the appreciation. We reviewed the Cochrane Handbook which in Chapter 10 discusses the influence of funding and conflicts of interest on the direction of results and conclusions of the SRs. Evidence is provided as to how ‘‘External funding has been found to be associated with publication independently of the statistical significance of the results’’ or ‘‘[...] studies sponsored by pharmaceutical companies were more likely to have outcomes favouring the sponsor than were studies with other sponsors’’. We have included the following refferences (# 10-21) to support the statement (page 8, lines 662-670):

Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. JAMA 2003;289:454-65.
Lexchin J, Bero LA, Djulbegovic B et al. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. BMJ 2003;326:1167-70.
Stelfox HT, Chua G, O’Rourke K et al. Conflict of interest in the debate over calcium-channel antagonists. N Engl J Med 1998;338:101-6.

Q6: Line 83-86. The importance of each is known to vary from one area of study to another. On the other hand, it has been quite established that the appraisal of RoB according to domains is a better approach than overall scores. See e.g., the reference cited above.

R6: We appreciate your valuable comments. As it seems that the sentence generates some kind of confusion, we have considered to remove it from the final version of the manuscript.

Q7: Line 97: I am not sure what historical articles are.

R7: This is an unfortunate expression that we have proceed to remove in the new manuscript.

Q8: Line 114: Why you did not include DARE?

R8: The reflection of the reviewer is very interesting. One of the limitations of our study is that the search was restricted to the cited databases given that our intention was to obtain a representative sample of SRs on treatments in psoriasis. Time and budget restraints require the authors to balance the thoroughness of the search with efficiency in use of time and funds. In this sense, the results of our systematic search are not only limited by the strategy used and by the number of databases consulted, as we did not review the gray literature since our intention was to evaluate published reviews. It is essential to consider this to avoid the risk of bias in the extraction of conclusions from the study. We have pointed out this deficit at 'Limitations and strenghts' subsection of the Discussion section (page 6, lines 473-477) as follow: ‘‘We did not seek SRs in grey literature databases, and, therefore, we cannot establish differences of methodological quality and risk of bias with respect to those that were examined.’’. Thank you very much for the appreciation.

Sections 2.3 and 2.4.
Q9: Firstly, in keeping with what I discuss in the results section of the abstract, I find confusing the titles of sections 2.3 and 2.4.

R9: As we have already commented, we have used the concepts of methodological quality and RoB that are included in the Cochrane Handbook. In relation to them, we have selected the tools that according to literature best evaluate them; AMSTAR for the evaluation of methodological quality, and ROBIS for risk assessment of bias. We have considered both tools although none of them are universally accepted. However, in order to clarify the confusion, we have modified 2.3 and 2.4 section titles to ‘‘Evaluation with AMSTAR instrument’’ (page 3, line 117) and ‘‘Assessment using ROBIS tool’’ (page 3, line 129), respectively. Thank you very much for your appreciation. We also change the titles of subbsection 3.3 and 3.4 of the Results section as follow: ‘‘Results using AMSTAR instrument’’ (page 4, line 211) and ‘‘Results using ROBIS tool’’ (page 4, line 228) , respectively.

Q10: (...) Secondly, Randomization of the order of the papers reviewed and especially of the order in which tools are applied seems, would be highly advisable.

R10: We consider of great interest this observation, as we did not consider randomizing the evaluations of the order of the papers nor the order in which the tools were used in the preparation of the PROSPERO protocol. We will include a sentence about this methodological limitation at the Discusion section of the manuscript as follows: ‘‘A limitation of this work is that we did not randomize the order in which the authors reviewed the papers or the order in which the evaluation was performed with both tools.’’ (page 6, lines 477-480).

Q11: Line 217-220: I am not sure how did you do it. Did you perform a pilot using some of the SR? Did you discard those SR from the final results?

R11: Thanks for the observation. We do not perform any pilot study using these tools. After raters independently evaluated each review, agreement among them was assessed in order to quantify the magnitude of differences. In those cases of disagreement an independent researcher (JR) was consulted to achieve a final consensus. All evaluated SRs were included in the final analysis.

Q12: Line 241: I think there is a mistake here. D1 is Study eligibility criteria.

R12: Thank you very much for point this out. We have solved this error in the new version of the manuscript (pg. 4, Ln 236).

Q13: Line 262. Figs. 2ad refer to Figs. 2a to 2d?

Q13: Yes, it does. We have changed ‘‘Figs. 2ad’’ by ‘‘Figs. 2a and 2d’’ in the new version (pg. 4, ln 257). Thank you very much for the observation.

Q14: Line 329: In keeping with what I mentioned before, I think it’s more appropriate to refer to the relationship between both tools, not between two nested concepts.

R14: Thank you for the helpful suggestion. We agree with the reviewer that it probably will be more appropriate to title this Results subsection as: ‘‘Correspondence between AMSTAR and ROBIS results’’, instead of ‘‘Relationships between methodological quality and risk of bias’’, as we have modified in the new version of the article (pg. 5, ln 321-322).

Q15: Line 353-358: This points to a problem of validity. What accuracy has to do with it?

R15: It is very interesting because there is a lot of discussion around this topic. As previously commented, we have adopted the Cochrane definition of methodological quality and RoB concepts. Our results provide evidence for this difference in the sense that a high methodological quality, measured with AMSTAR, does not imply a low RoB, assessed using ROBIS. We agree with the reviewer. Indeed, the mentioned statement does not reflect our message properly. Actually, we consider that found discrepancies between AMSTAR vs ROBIS classification it is not a question of accuracy. We have modified the above mentioned paragraph as follows: ‘‘Overall, our results suggest that methodological quality only explains a proportion of the bias risk of SRs, as we observed that most of reviews classified as high and moderate methodological quality by AMSTAR were also considered as displaying a high risk of bias using ROBIS.’’ (pg. 5, ln 348-353). We appreciate this comment from the reviewer.

Q16: Line 361-364: I don’t understand how this statement derives from previous thoughts or from the results.

R16: This statement (‘‘[...] similar to the evaluation of primary studies, it is possible to carry out a SR following the highest methodological standards and still having a high risk of bias’’) derives from our results, as we observed that most of reviews classified as high and moderate methodological quality by AMSTAR were also considered as displaying a high risk of bias using ROBIS, and it is in line with the statement at the Cochrane Handbook related to primary studies: ‘‘[...] A study may be performed to the highest possible standards yet still have an important risk of bias. For example, in many situations it is impractical or impossible to blind participants or study personnel to intervention group. It is inappropriately judgemental to describe all such studies as of ‘low quality’, but that does not mean they are free of bias resulting from knowledge of intervention status [Chapter 8 of the Cochrane Handbook]’’.

Q17: Line 382-389. I don’t understand your point. I don’t see the relationship between the lack of information (reporting problems) in systematic reviews.

R17: Thank you very much for this observation. The loss of information is one of the great problems of evidence-based medicine. As SRs summarize the results of clinical trials, that a great proportion of primary studies report missing data finally affects the quality of the reviews including them. In our study, we found a high percentage of SRs, similar to the percentage communicated for primary studies. The aim of this comment was to highlight that the impact of missing data concerning SRs is of similar magnitude to primary studies.

Q18: Section 4.4. Weighting for different items is a problematic solution that might only make sense taking into account the context of the SR, which makes this weighting even more complex.

R18: We agree with the reviewer's assessment. We used the PCA to explore the mutidimensional nature of our dataset and looking for new subgroups inside of those defined according to AMSTAR levels. Our objective was to use PCA as an exploratory analysis, not as an approach to get the weights to improve AMSTAR tool.

Q19: Section 5. If AMSTAR have all those limitations, why should be complementary to Robis? If so, please provide guidance on how to combine them.

R19: Both ROBIS and AMSTAR evaluate the methodological quality of systematic reviews. ROBIS also takes into account the methodological compliance in the assessment the RoB. However, as described in the discussion, there are differences in the construction of both tools, mainly as regards of financing, conflict of interest, and lost data. Thus, there are aspects related to methodological quality that are not taken into account if only one of these tools is used. The compliance with both tools increase the likelihood of methodological quality. We have not considered a better way to combine them than to perform both assessment and interpret their results in a combined 3 x 3 way: high, low, uncertain risk of bias and high, medium, low methodological quality, respectively.

Q20: Figure 2. I found the figures, footnotes and the corresponding text in Section 3.5 confusing.

R20: We wholeheartedly agree with the reviewer’s comment, and have revised this confusing part of the Results section of the document as follows:
New version of Fig. 2 caption: (page 11)
‘‘Fig. 2. Scale reduction and high-dimensional visualization of AMSTAR and ROBIS results using principal component analysis (PCA).
PCA was performed using the 11 AMSTAR item subscores or the responses to 21 signaling questions of ROBIS per article. Fig 2a and 2d display PC1-PC2 projections of every review, using different shapes to identify the risk of bias or the level of methodological quality. Fig. 2b and 2e show the contribution of each variable on PC1 and PC2. Fig. 2c and 2f display the contribution of each review to PC1 and PC2. A color gradient represent the magnitude of variables and review contributions. Fig. 2g-j show Radviz data visualization of ROBIS phase 2 domains judgments and response profiles to signaling questions. Points represent reviews and are colored with respect to the risk of bias (low-turquoise, high-coral) or methodological quality classification (high-blue, moderate-green, low-red).’’

New version text paragraph related to Fig. 2 (page 4, lines 252-300)

‘‘We used PCA to convert vectors of 11 AMSTAR item subscores and answers to the 21 ROBIS signalling questions per article into two sets of values of linearly uncorrelated variables called principal components (PCs), or projections to anchored domains or questions, respectively. Figs. 2a and 2d show two PCA scatterplots that comprise PC1 and PC2 projections of all included reviews. Overlapping was more evident between high or low risk of bias reviews as compared to reviews demonstrating high vs. moderate, or moderate vs. low methodological quality. A scree plot of AMSTAR-based PCA data showed that the first component (PC1) explained 30\% of variance, and that components PC1, PC2, PC3, and PC4 explained more than 50\% of this variability (Fig. S2a). When considering ROBIS-based PCA data, the scree plot displayed a different result: component PC1 explained more than 45\% of variance, while each of the following components contributed individually to less than 5-10\% to this variability (Fig. S2b).
We further analyzed how each item or question contributed to explain the observed variability. For AMSTAR-based PCA, QA8 ('Was the scientific quality of the included studied used appropriately in formulating conclusions?') and QA7 ('Was the scientific quality of the included studies assessed and documented?') were the items that contributed the most to discriminate between reviews, while QA6 (Were the characteristics of the included studies provided?') contributed the least (Fig. 2b). In the case of ROBIS-based PCA, QR5 ('Were efforts made to minimize error in risk of bias assessment?') and QR34 ('Was risk of bias [or methodological quality] formally assessed using appropriate criteria?') were the signalling questions that most contributed to explain variability of risk of bias among reviews (Fig. 2e).
Fig. 2g-h represent Radviz plots showing how the 21 signalling questions separate high and low risk of bias reviews apart. While low risk reviews clustered together in the center of the circle, high risk reviews were more sparced and some of them were overlaping with the formers. This fact shows that no perfect separation between low and high risk reviews is obtained when response to all signaling questions are considered. In Fig. 2i-j reviews are tagged by colors based on methodological classification by AMSTAR.

Q21: Figure 3b. Please make more clear what No/Yes categories are.

R21: We agree with the reviewer's comment. Figure caption is not clear for 3b plot. In the new version of the manuscript, we have clarified it as follow: ‘‘[...] This plot shows frequency distributions responses ('no' or 'yes') of AMSTAR per item subscores comparing reviews by risk of bias using ROBIS tool ('high' or 'low')’’ (page 12). A brief modification of 3b plot legend has been added too.

Q22: Figures 3d and e: It appears to me that both figures show similar profiles while one correspond to High RoB and the other to low RoB. I find it really surprising but I didn’t find it discussed.

R22: It is a very interesting appreciation on the part of the reviewer. These figures show ROBIS phase 2-domains compliance comparing high vs low RoB reviews. Both profiles

are quite similar, but the frequencies of 'low' and 'probably no' responses to the signaling questions are lower for the low RoB reviews. QR33 is the question that variates the least between both groups. As suggested, a comment has been included in the new version of the paper as follow: ‘‘Both profiles are quite similar, but the frequencies of 'low' and 'probably no' responses to the signaling questions are lower for low as compared with high RoB reviews.‘‘ (pg 5, ln 335-338).

Reviewer 2
Q1: The authors would like to suggest on what basis the AMSTAR evaluation results for SR are evaluated as high, moderate, low in the manuscript.

R1: Detailed information for the system of rating is presented in Supplementary materials and methods. Specifically, the highest possible AMSTAR score is 11. Review quality was classified by AMSTAR score following quality levels with similar cutoff points used by most of studies [for low (0-4), moderate (5-8), and high methodological quality (9-11) respectively].

Q2:I think if the same evaluator made an assessment of AMSTAR and robis, the results would have been more valid. I believe this is the limit of research.

R2: We appreciated this comment very much. Only one rater (FG-G) did the evaluation of the reviews using both tools, and their results were compared with raters that only used AMSTAR (JG-M) or ROBIS (MA-L) instruments. Although their results were compared in pairs and discrepancies were discussed with a fourth rater, there is a risk that this fact will affect the validity of our results, so we will take into account this methodological limitation in the Discussion section of the manuscript as follow: ‘‘Finally, only one of threes raters carried out the evaluations both with AMSTAR and ROBIS tools. Although their results were compared in pairs and discrepancies were discussed with a fourth rater, there is a risk that this issue will affect the validity of our results.’’. (page 6, lines 430-435).

Q3: In the conclusion, ‘‘AMSTAR and ROBIS tools may be considered as complementary instruments to assess the quality of SRs’’. For example, the ratio of high SR quality in current manuscripts is 22.3\% for AMSTAR (75.5\% for above moderate) and 14\% for ROBIS. How do you interpret these results?

R3: We have employed the concepts of risk of bias (RoB) and methodological quality included in the Cochrane Handbook. In this sense ‘‘The ‘assessment of methodological quality’ suggests an investigation of the extent to which study authors conducted their research to the highest possible standards. This would be different from the RoB, or risk a systematic error in conducting the study. It is therefore plausible that the results of AMSTAR do

not correspond to those of ROBIS. Even more so because in ROBIS does not only evaluate methodological compliance but the authors' interpretation of the limitations of their work. In our opinion, these results support that there are differences between the concepts of methodological quality and risk of bias. The methodological quality explains only part of the risk of bias of SRs.

Reviewer 3

It is interesting study. I have several questions and comments.

Q1: Page 3. line 141 and 149. Sentence "Phase 1 assess ... " seems to be repeated.

R2: This must be an error. Thank you very much for pointing this out. We will delete the second sentence as suggested.

Q2:It is good to show AMSTAR and ROBIS checklist as supplementary table. I think it would be better comparing those items using table rather than description.

R2: Thanks for the observation. Due to the space constraints in the main text, we have considered including AMSTAR and ROBIS checklist as tables (S4 and S5) in the supplementary material.

Q3: Conclusion. ‘‘AMSTAR and ROBIS tools may be considered as complementary ...’’ this sentence seems to be vague. you can give more clear advice for future researchers, such as "we recommend to use both tool when conducting quality assessment...".

R3: We agree with reviewer. In the new version of the paper, the statement "We recommend to use both AMSTAR and ROBIS tools when conducting quality assessment of SRs, as they may be considered as complementary instruments’’ was added in the Conclusion subsection (page 7, lines 564-567).

We would like to thank all reviewers for their insightful comments, and for their very and careful review of our paper. After completion of the suggested edits, the revised manuscript has tremendously benefitted from an improvement in the overall clarity of the presentation.

Code scripts.

Stata code for paper #1

```
*David Epstein, University of Granada, 2016
*Network metanalysis of pasi score
/*
Web addresses to install meta analysis commands:
net from http://www.mtm.uoi.gr
net install network_graphs, replace
*For the metan command:
net install sbe24_3,from(http://www.stata-journal.com/software/sj9-2) replace
*For the mvmeta command:
net install mvmeta,from(http://www.mrc-bsu.cam.ac.uk/IW_Stata/meta) replace
*For the metareg command:
net install sbe23_1,from(http://www.stata-journal.com/software/sj8-4) replace
*/
cd "C:\Users\workfiles"
set more off
cap log close
local out=0
foreach outcome in PASI90  PASI75  {
local out=`out'+1
log  using pasi`out',replace
clear
import excel "database_NMA_BJD.xls", firstrow
describe
drop I J K
gen d = `outcome'
replace intervention = "0_Placebo" if intervention=="Placebo"
replace doses = subinstr(doses, " mg", "mg", .)
gen trt_dose = intervention + " " + doses
drop if d==.
levelsof trt_dose, local(tlabel)
```

```
display r(levels)
encode trt_dose, gen(trt_code)
list ID trt_code n d,  sepby(ID)
keep ID  n d  trt_code
*How many treatments?
summ trt_code
local n_trt = r(max)
*PREPARATION OF DATASET IN WIDE FORMAT FOR MVMETA
reshape wide n d , i(ID) j(trt_code)
*The Stata Journal (2011) 11, Number 2, pp. 255-270
*calculate variances and covariances of log odds ratios
replace d1= 0.001 if d1==.
replace n1= 0.1 if n1==.
forvalues trt = 1(1)`n_trt' {
      if `trt'==1 continue  // continue means go to next forvalues
      gen y`trt' = log(d`trt'/(n`trt'-d`trt')) - log(d1/(n1-d1))
      gen S`trt'`trt' = 1/d`trt' + 1/(n`trt'-d`trt') + 1/d1 + 1/(n1-d1)
      forvalues trt2 = 1(1)`n_trt' {
            if `trt2'==1 continue
            if `trt2' > `trt' {
gen S`trt'`trt2' = 1/d1 + 1/(n1-d1) if !mi(d`trt') & !
mi(d`trt2')
            }
} }
format y* S* %6.2g
sort ID
mat P = I(`n_trt' -1) + J(`n_trt' -1,`n_trt' -1,1) //matrix size is number of
treatments minus 1
mat l P
*Note that we assume a common heterogeneity estimate for all comparisons (with
the bscov()option).
mvmeta y S, bscov(prop P ) pbest(max in 1, zero all gen(prob) cum line predict
saving(f1) replace)
estimates save mvmeta_19_09_`out', replace
save mvmeta`out', replace
use mvmeta`out',clear
estimates use mvmeta_19_09_`out'
*MIXED COMPARISONS
tempname memhold
tempfile results
postfile `memhold' str20 t1 str20 t2 or_mix seor using `results'
local i=1
foreach l of local tlabel {
      if strmatch(rtrim("`l'"),"0_Placebo") continue
```

```
      local i = `i'+1
      local j=1
      foreach m of local tlabel {
            if strmatch(rtrim("`m'"),"0_Placebo") continue
            local j = `j' + 1
            if `i'!=`j' {
} }
}
lincom _b[y`j':_cons] - _b[y`i':_cons]
post `memhold' ("`m'") ("`l'") (r(estimate)) (r(se))
postclose `memhold'
use `results',clear
gen cilow_mix = exp(or_mix - 1.96*seor)
gen cihigh_mix = exp(or_mix + 1.96*seor)
replace or_mix = exp(or_mix)
sort t1
list t1 t2 or_mix cilow_mix cihigh_mix
keep t1 t2 or_mix cilow_mix cihigh_mix
order t1 t2 or_mix cilow_mix cihigh_mix
save  mixed`out',replace
export excel mixed`out', firstrow(varlabels) replace
use mvmeta`out', clear
estimates use mvmeta_19_09_`out'
*SUCRA PLOT
sucra prob*, mvmetaresults  lab(`tlabel')
graph export sucra`out'.tif,replace
graph export sucra`out'.ps,replace
/*
Note that we assume a common heterogeneity estimate for all comparisons (with
the bscov()option).
A predictive interval plot could be produced by typing:
*intervalplot,mvmeta pred eform  null(1)
log close
}

================================
*PREPARE DATA FOR NETWORKPLOT
clear
local out = 0
foreach outcome in PASI90  PASI75  {
local out = `out' + 1
estimates use mvmeta_19_09_`out'  //suffix 1 is pasi 90
matrix Sigma = e(Sigma)
global tau = Sigma[1,1] //between study variance from mvmeta
```

```
di `tau'
set more off
cap log close
log using preparedata, replace
clear
import excel "database_NMA_BJD.xls", firstrow
describe
replace intervention = "0_Placebo" if intervention=="Placebo"
rename `outcome' d
drop if d==.
replace doses = subinstr(doses, " mg", "mg", .)
gen trt_dose = intervention + " " + doses
levelsof trt_dose, local(tlabel)
display r(levels)
encode trt_dose, gen(T)
list ID T n d,  sepby(ID)
keep ID T n d
rename n n1
rename d d1
rename T T1
save data_1, replace
rename n1 n2  //create copy of dataset with different var names
rename d1 d2
rename T1 T2
save data_2, replace
use data_1
joinby ID using data_2  // crosses all combinations
sort ID T1 T2
drop if T2==T1
drop if T2>T1
rename T1 t1
rename T2 t2
rename d1 r1  //outcome variable is named r in networkplot help file
rename d2 r2
save pasi_pairwise.dta,replace
log close
*metanalysis of pasi score
cap log close
log using pasi_pairwise`out',replace
use "pasi_pairwise.dta", clear
/*
To produce the network plot of Figure 1 where the width of the edges is
proportional to the mean control group risk for all comparisons versus placebo,
first calculate the control group risk for studies including the placebo:
```

```
*/
gen cgr=r1/n1 if t1==1
replace cgr=r2/n2 if t2==1
*gen contrast = t1 + " v " + t2
*tab contrast
replace r1=0.5 if r1==0
replace n1=n1+0.5 if r1==0.5
replace r2=0.5 if r2==0
replace n2=n2+0.5 if r2==0.5
gen logOR = log(r2/(n2-r2)) - log(r1/(n1-r1))
gen selogOR = sqrt(1/r1 + 1/(n1-r1) + 1/r2 + 1/(n2-r2))
save pasi_pairwise.dta,replace
*Aurelio Tobias's code for indirect comparisons
keep if t1==1 | t2==1 //placebo comparisons
levelsof t1, local(t1label)
display r(levels)
tabulate t1, generate (x)
sort  t1
list ID t1 t2, sepby(t2) noobs
metareg logOR x*, wsse(selogOR) mm z eform noconstant
*Indirect comparisons via placebo
tempname memhold
tempfile results
postfile `memhold' str20 t1 str20 t2 or_ind seor using `results'
local i=0
foreach l of local t1label {
      *if strmatch(rtrim("`l'"),"0_Placebo") continue
      local i = `i'+1
      di `i'
      local lnum = real("`l'")
      di `lnum'
      local lvar `: word `lnum' of `tlabel''
      local j=0
      foreach m of local t1label {
} }
*if strmatch(rtrim("`m'"),"0_Placebo") continue
local j = `j' + 1
local mnum = real("`m'")
local mvar `: word `mnum' of `tlabel''
di "`mvar'" "`lvar'"
lincom x`i' - x`j'
post `memhold' ("`mvar'") ("`lvar'") (r(estimate)) (r(se))
postclose `memhold'
use `results',clear
```

```
gen cilow_ind = exp(or_ind - 1.96*seor)
gen cihigh_ind = exp(or_ind + 1.96*seor)
replace or_ind = exp(or_ind)
sort t1
list t1 t2 or_ind cilow_ind cihigh_ind
keep t1 t2 or_ind cilow_ind cihigh_ind
save  indir`out',replace

********Direct treatment effects
tempname memhold
tempfile results
postfile `memhold' str20 t1 str20 t2 or_dir seor using `results'
local i=0
foreach l of local tlabel {
local i = `i'+1
 local j=0
      foreach m of local tlabel {
            use pasi_pairwise.dta,clear
            local j = `j'+1
            list t1 t2 logOR selogOR if t1==`j' & t2==`i'
            keep  if t1==`j' & t2==`i'
            if _N==1 {  //only one trial so no metanalysis
                  local a = -(logOR[1])  // inverse
                  local b = selogOR[1]
                  post `memhold' ("`m'") ("`l'") (`a') (`b')
                  }
            if _N>1 {  //results of metanalysis
                  metareg logOR, wsse(selogOR) mm z eform
                  matrix A = e(b)
                  local a = -(A[1,1])  //inverse
                  matrix B = e(V) //variance
                  local b = sqrt(B[1,1]) //se
                  post `memhold' ("`m'") ("`l'") (`a') (`b')
                  }
} }
postclose `memhold'
use `results',clear
gen cilow_dir = exp(or_dir - 1.96*seor)
gen cihigh_dir = exp(or_dir + 1.96*seor)
replace or_dir = exp(or_dir)
sort t1
list t1 t2 or_dir cilow_dir cihigh_dir
keep t1 t2 or_dir cilow_dir cihigh_dir
save  direct`out',replace
```

```
use pasi_pairwise.dta,clear
decode t1 , generate(tr1)
replace tr1 = "Placebo" if tr1== "0_Placebo"
decode t2 , generate(tr2)
replace tr2 = "Placebo" if tr2== "0_Placebo"
gen comp = tr1 + " v " + tr2
sort comp ID
replace logOR = -logOR  //inverse
metan logOR selogOR, random effect(OR) eform by(comp) nooverall /*
*/   lcols(ID  ) nowt
graph export direct_`out'.tif, replace
*Inconsistency plot
use pasi_pairwise.dta,clear
ifplot logOR selogOR t1 t2 ID,
log close
}
*Network plot
networkplot t1 t2, edgew(cgr*50 mean) saving(network_2,replace) lab(`tlabel')
graph export network.tif, replace width(6000)
graph export network.ps, replace
graph save networkplot, replace
```

R code for paper #2

```
#' ---
#' title: "Methodological appraisal of systematic reviews and
#' meta-analyses on psoriasis: role of funding sources,
#' conflict of interest and pressure to publish"
#' author: "Juan Ruano"
#' date: "30 Sep 2016"
#' institutions: Department of Dermatology,
#' IMIBIC/Reina Sofia University Hospital/University of Cordoba, Cordoba, Spain
#' analyse: 01_Statistical analyses: descriptive, PCA, heatmaps, and regression model
#' ---
#'
#' R version 3.3.1 (2016-06-21)
#' Platform: x86_64-apple-darwin13.4.0 (64-bit)
#' Running under: OS X 10.9.5 (Mavericks)


########  R packages ----------------

library(ggplot2)
library(lubridate)
library(ggfortify)
library(psych)
library(MASS)
library(nFactors)
library(tabplot)
library(plyr)

########  environment setting ----------------
# For RStudio only

setwd(dirname(rstudioapi::callFun("getActiveDocumentContext")$path))


######## Read .csv files --------------------------------

file1<-read.csv2("AMSTARfinal_metadatos_articulos_con_topics.csv",
                 sep = ";",
                 dec= ".",
                 stringsAsFactors = TRUE,
                 header = TRUE)
DB1<-as.data.frame(file1)
```

```
########  Tiding dataset ------------------------------
DB1$Total.Cites <- as.numeric(DB1$Total.Cites)
DB1$Cited.Half.life <- as.numeric(DB1$Cited.Half.life)
DB1$Citing.Half.life <- as.numeric(DB1$Citing.Half.life)
DB1$Normalized.Eigenfactor <- as.numeric(DB1$Normalized.Eigenfactor)
DB1$num_institutions <- as.numeric(DB1$num_institutions)
DB1$conflict_of_interest <- as.numeric(DB1$conflict_of_interest)

DB1$funding_academic <- as.factor(DB1$funding_academic)
DB1$funding_industry <- as.factor(DB1$funding_industry)
DB1$funding <- as.factor(DB1$funding)

DB1$metaanalysis_included <- factor(DB1$metaanalysis_included)
DB1$AMSTAR_levels <- factor(DB1$AMSTAR_levels,
levels=c("low_quality", "moderate_quality", "high_quality"))
DB1$AMSTAR_levels_2 <- factor(DB1$AMSTAR_levels_2,
                    levels=c("low_quality", "moderate_quality", "high_quality"))
DB1$SJR.Best.Quartile <- factor(DB1$SJR.Best.Quartile,
levels=c("Q1", "Q2", "Q3", "Q4"))
DB1$Country <- factor(DB1$Country,
levels=c("Brazil", "Canada", "China", "Germany", "India",
                    "Ireland", "Netherlands", "Sweden",
                    "Switzerland", "United Kingdom"))
DB1$topic <- factor(DB1$topic,
levels=c("Comorbidities",
                    "Economic analyses",
                    "Pathogeny",
                    "Treatment"))
DB1$year <- factor(DB1$year,
levels=c("1999",
                    "2000",
                    "2001",
                    "2002",
                    "2003",
                    "2004",
                    "2005",
                    "2006",
                    "2007",
                    2008",
                    "2009",
                    "2010",
                    "2011",
```

```
                    "2012",
                    "2013",
                    "2014",
                    "2015",
                    "2016"))

levels(DB1$funding_industry)      <- c("No","Funded by Pharma")
levels(DB1$funding_academic)      <- c("No","Funded by Academic")
levels(DB1$metaanalysis_included) <- c("No","Metanalisis included")
levels(DB1$clinical_trials)       <- c("No","Analyses of clinical_trials")
levels(DB1$observational_studies) <- c("No","Analyses of observational studies")
levels(DB1$systematic_reviews)    <- c("No","Analyses of systematic reviews")
levels(DB1$economic_studies)      <- c("No","Analyses of economic studies")
levels(DB1$AMSTAR_levels_2)       <- c("low quality", "moderate quality", "high quality")
levels(DB1$AMSTAR_levels_2)       <- factor(DB1$AMSTAR_levels_2,
levels=c("low quality",
                                        "moderate quality",
                                        "high quality"))

DB1$topic_treatment <- mapvalues(DB1$topic,
from = c("Comorbidities",
                         "Economic analyses",
                         "Pathogeny",
                         "Treatment"),
                         to = c("0", "0", "0", "Treatment"))
DB1$topic_Pathogeny <- mapvalues(DB1$topic,
from = c("Comorbidities",
                         "Economic analyses",
                         "Pathogeny",
                         "Treatment"),
                         to = c("0", "0", "Pathogeny", "0"))
DB1$topic_Economic_analysis <- mapvalues(DB1$topic,
from = c("Comorbidities",
                         "Economic analyses",
                         "Pathogeny",
                         "Treatment"),
                         to = c("0", "Economic analyses", "0", "0"))
DB1$topic_Comorbidities <- mapvalues(DB1$topic,
from = c("Comorbidities",
                         "Economic analyses",
                         "Pathogeny",
                         "Treatment"),
                         to = c("Comorbidities", "0", "0", "0"))
```

```
######## plotting heat maps --------------------------------
####  R packages ----------------

source("https://bioconductor.org/biocLite.R")
biocLite()
biocLite("ComplexHeatmap")
library(ComplexHeatmap)
library(circlize)
library(dendsort)
library(dendextend)


#### prepare matrix  ----------------
DB_heat <- na.omit(DB1[c(3,26,69,66,63,62,72,73,75,77,76,31,39,33,23,24,7,19:22, 56,57)])
DB_heat <- DB_heat[order(DB_heat$sort_AMSTAR_levels),]
heatDB <- as.matrix(DB_heat[c(1,18:21)])

topic_df <- (DB_heat[c(1,2)])
topic <- as.matrix(topic_df)

number_authors_df <- (DB_heat[c(1,22)])
number_authors <- as.matrix(number_authors_df)

number_institutions_df <- (DB_heat[c(1,23)])
number_institutions<- as.matrix(number_institutions_df)

article_page_count_df <- (DB_heat[c(1,3)])
article_page_count <- as.matrix(article_page_count_df)

conflict_of_interest_df <- (DB_heat[c(1,4)])
conflict_of_interest <- as.matrix(conflict_of_interest_df)

funding_industry_df<- (DB_heat[c(1,5)])
funding_industry                 <- as.matrix(funding_industry_df)

funding_academic_df <- (DB_heat[c(1,6)])
funding_academic <- as.matrix(funding_academic_df)

clinical_trials_df <- (DB_heat[c(1,7)])
clinical_trials <- as.matrix(clinical_trials_df)

observational_studies_df <- (DB_heat[c(1,8)])
observational_studies<- as.matrix(observational_studies_df)
```

```
systematic_reviews_df <- (DB_heat[c(1,9)])
systematic_reviews<- as.matrix(systematic_reviews_df)

metaanalyses_included_df<- (DB_heat[c(1,10)])
metaanalyses_included <- as.matrix(metaanalyses_included_df)

economic_studies_df<- (DB_heat[c(1,11)])
economic_studies <- as.matrix(economic_studies_df)

journal_impact_factor_df <- (DB_heat[c(1,12)])
journal_impact_factor <- as.matrix(journal_impact_factor_df)

article_influence_score_df <- (DB_heat[c(1,13)])
article_influence_score <- as.matrix(article_influence_score_df)

five_year_impact_factor_df <- (DB_heat[c(1,14)])
five_year_impact_factor<- as.matrix(five_year_impact_factor_df)

google_cytes_df<- (DB_heat[c(1,15)])
google_cytes <- as.matrix(google_cytes_df)

web_of_science_df <- (DB_heat[c(1,16)])
web_of_science  <- as.matrix(web_of_science_df)

type_df <- (DB_heat[c(1,17)])
type <- as.matrix(type_df)


#### individual heat maps ----------------

topicAreaPlot <- Heatmap(topic[,2],
                   name="Topic area of research",
                   show_row_names = FALSE,
                   cluster_rows = FALSE,
                   width = unit(5, "mm"),
                   show_heatmap_legend = TRUE,
                   col = c("navyblue","lightgreen", "chocolate4", "darkolivegreen"))

numAuthorsPlot <- Heatmap(number_authors[,2],
                   name="Number of authors",
                   show_row_names = FALSE,
                   cluster_rows = FALSE,
                   width = unit(5, "mm"),
                   heatmap_legend_param = list(color_bar = "continuous"),
```

```
                    show_heatmap_legend = FALSE,
                    c("white", "steelblue4"))

numInstitutionsPlot <- Heatmap(number_institutions[,2],
                     name="Number of institutions",
                     show_row_names = FALSE,
                     cluster_rows = FALSE,
                     width = unit(5, "mm"),
                     heatmap_legend_param = list(color_bar = "continuous"),
                     show_heatmap_legend = FALSE,
                     c("white", "royalblue1"))

articlePageCountPlot<- Heatmap(log(article_page_count[,2],2),
                     name="Article page count (log2)",
                     show_row_names = FALSE,
                     cluster_rows = FALSE,
                     width = unit(5, "mm"),
                     show_heatmap_legend = FALSE,
                     col = colorRamp2(c(2, 8),
                     c("white", "forestgreen")))

numAuthorsConflictInterestPlot   <- Heatmap(conflict_of_interest[,2],
                     name="Number of authors with conflict of interest",
                     show_row_names = FALSE,
                     cluster_rows = FALSE,
                     width = unit(5, "mm"),
                     #col = colorRamp2(c(0, 15),
                     show_heatmap_legend = FALSE,
                     heatmap_legend_param = list(color_bar = "continuous"),
                     c("white", "aquamarine4"))

fundingPharmaPlot<- Heatmap(funding_industry[,2],
                     name="Funding by pharmaceutical companies",
                     show_row_names = FALSE,
                     cluster_rows = FALSE,
                     width = unit(5, "mm"),
                     show_heatmap_legend = FALSE,
                     col = c("aquamarine3", "white"))

fundingAcademicPlot <- Heatmap(funding_academic[,2],
                     name="Funding by academic institutions",
                     show_row_names = FALSE,
                     cluster_rows = FALSE,
                     width = unit(5, "mm"),
```

```
                    show_heatmap_legend = FALSE,
                    col = c("firebrick1", "white"))

clinicalTrialsPlot <- Heatmap(clinical_trials[,2],
                    name="Clinical trials analyses",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
                    show_heatmap_legend = FALSE,
                    col = c("white", "gray39"))

observationalStudiesPlot <- Heatmap(observational_studies[,2],
                    name="Observational studies analyses",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
                    show_heatmap_legend = FALSE,
                    col = c("white", "seashell4"))

srPlot<- Heatmap(systematic_reviews[,2],
                    name="Systematic reviews analyses",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
                    show_heatmap_legend = FALSE,
                    col = c("white", "cadetblue4"))

metaanalysesIncludedPlot<- Heatmap(metaanalyses_included[,2],
                    name="Meta-analyses included",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
                    show_heatmap_legend = FALSE,
                    col = c("cadetblue3", "white"))

economicStudiesPlot  <- Heatmap(economic_studies[,2],
                    name="Economic analyses included",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
                    show_heatmap_legend = FALSE,
                    col = c("white", "darkorange1"))

journalImpactFactorPlot<- Heatmap(log(journal_impact_factor[,2], 2),
```

```
                    name="Journal impact factor (log2)",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
                    #col = colorRamp2(c(-3, 4),
                    heatmap_legend_param = list(color_bar = "continuous"),
                    show_heatmap_legend = FALSE,
                    c("white", "tan4"))

articleInfluenceScorePlot<- Heatmap(log(article_influence_score[,2],2),
                    name="Article influence score (log2)",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
                    #col = colorRamp2(c(-3, 4),
                    heatmap_legend_param = list(color_bar = "continuous"),
                    show_heatmap_legend = FALSE,
                    c("white", "tan3"))

fiveYearImpactFactorPlot<- Heatmap(log(five_year_impact_factor[,2],2),
                    name="5-year impact factor (log2)",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
                    #col = colorRamp2(c(-0.3, 5),
                    heatmap_legend_param = list(color_bar = "continuous"),
                    show_heatmap_legend = FALSE,
                    c("white", "tan1"))

citesInGoogleScholarPlot<- Heatmap(log(google_cytes[,2]+0.1,2),
                    name="Cites in Google Scholar (log2)",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
                    #col = colorRamp2(c(0, 100),
                    heatmap_legend_param = list(color_bar = "continuous"),
                    show_heatmap_legend = FALSE,
                    c("white", "orange"))

citesInWoSPlot <- Heatmap(log(web_of_science[,2]+0.1,2),
                    name="Cites in Web of Science (log2)",
                    show_row_names = FALSE,
                    cluster_rows = FALSE,
                    width = unit(5, "mm"),
```

```
                    #col = colorRamp2(c(0, 100),
                    heatmap_legend_param = list(color_bar = "continuous"),
                    show_heatmap_legend = FALSE,
                    c("white", "coral"))

amstarLevelsPlot <- Heatmap(type[,2],
                    name="AMSTAR levels",
                    show_row_names = FALSE,
                    width = unit(5, "mm"),
                    show_heatmap_legend = TRUE,
                    col = c("steelblue3", "red", "lightgreen"))

dend <- hclust(dist(heatDB[,2:5]))
dend <- color_branches(dend, k = 6)

pcaFactors<- Heatmap(heatDB[,2:5],
                    cluster_columns = FALSE,
                    cluster_rows = TRUE,
                    show_row_names = FALSE,
                    width = unit(20, "mm"),
                    km = 6,
                    row_order = 4:220,
                    show_heatmap_legend = FALSE,
                    show_column_names = TRUE)

######## plotting all heat maps together ----------------

map_list<-(pcaFactors+
  amstarLevelsPlot +
  topicAreaPlot +
  articlePageCountPlot +
  numAuthorsPlot +
  numInstitutionsPlot +
  numAuthorsConflictInterestPlot +
  fundingPharmaPlot +
  fundingAcademicPlot +
  clinicalTrialsPlot +
  observationalStudiesPlot +
  srPlot +
  metaanalysesIncludedPlot +
  economicStudiesPlot +
  journalImpactFactorPlot +
  articleInfluenceScorePlot +
  fiveYearImpactFactorPlot +
```

```
  citesInGoogleScholarPlot +
  citesInWoSPlot)

postscript("allHeatMaps.eps",
           width = 2000,
           height = 1000)
draw(map_list)
graphics.off()

tiff("allHeatMaps.tiff",
     width = 2000,
     height = 1000,
     pointsize = 1/300,
     units = 'in',
     res = 300)
draw(map_list)
dev.off()

######## get the items by cluster -----------------

itemsByCluster<-cutree(dend, k = 6)
write.table(itemsByCluster,
file="itemsByCluster.csv",
        quote=F, sep="\t",
        col.names=NA, append=T)


#' ---
#' title: "Systematic reviews and meta-analyses on psoriasis: role of funding sources,
#' conflict of interest, and bibliometric indices as predictors of methodological quality"
#' author: "Juan Ruano"
#' date: "30 Sep 2016"
#' institutions: Department of Dermatology, IMIBIC/Reina Sofia University Hospital/University
of Cordoba,
#' Cordoba, Spain
#' analyse: 01_Statistical analyses: descriptive, PCA, heatmaps, and regression model
#' ---
#'
#' R version 3.3.1 (2016-06-21)
#' Platform: x86_64-apple-darwin13.4.0 (64-bit)
#' Running under: OS X 10.9.5 (Mavericks)


################  regression model, cross-validation test and probability plots  ---------
```

```
########  R packages ----------------

library(ggplot2)
library(lubridate)
library(ggfortify)
library(psych)
library(MASS)
library(nFactors)
library(tabplot)
library(plyr)

require(foreign)
require(ggplot2)
require(MASS)
require(Hmisc)
require(reshape2)
library(rms)


########  read .csv files ----------------

file1<-read.csv2("AMSTARfinal_metadatos_articulos_con_topics.csv",
                 sep = ";",
                 dec= ".",
                 stringsAsFactors = TRUE,
                 header = TRUE)
DB1<-as.data.frame(file1)


####  tiding dataset ----------------

DB1$Total.Cites                  <- as.numeric(DB1$Total.Cites)
DB1$Cited.Half.life              <- as.numeric(DB1$Cited.Half.life)
DB1$Citing.Half.life             <- as.numeric(DB1$Citing.Half.life)
DB1$Normalized.Eigenfactor       <- as.numeric(DB1$Normalized.Eigenfactor)
DB1$num_institutions             <- as.numeric(DB1$num_institutions)
DB1$conflict_of_interest         <- as.numeric(DB1$conflict_of_interest)

DB1$funding_academic             <- as.factor(DB1$funding_academic)
DB1$funding_industry             <- as.factor(DB1$funding_industry)
DB1$funding                      <- as.factor(DB1$funding)

DB1$metaanalysis_included        <- factor(DB1$metaanalysis_included)
DB1$AMSTAR_levels                <- factor(DB1$AMSTAR_levels,
```

```
levels=c("low_quality",
                                  "moderate_quality",
                                  "high_quality"))
DB1$AMSTAR_levels_2              <- factor(DB1$AMSTAR_levels_2,
levels=c("low_quality",
                                  "moderate_quality",
                                  "high_quality"))
DB1$SJR.Best.Quartile            <- factor(DB1$SJR.Best.Quartile,
levels=c("Q1", "Q2", "Q3", "Q4"))
DB1$Country                      <- factor(DB1$Country,
levels=c("Brazil", "Canada", "China", "Germany", "India",
                                  "Ireland", "Netherlands", "Sweden", "Switzerland",
"United Kingdom"))
DB1$topic                        <- factor(DB1$topic,
levels=c("Comorbidities",
                                  "Economic analysis",
                                  "Pathogeny",
                                  "Treatment"))
DB1$year                         <- factor(DB1$year,
levels=c("1999",
"2000",
                                  "2001",
                                  "2002",
                                  "2003",
                                  "2004",
                                  "2005",
                                  "2006",
                                  "2007",
                                  "2008",
                                  "2009",
                                  "2010",
                                  "2011",
                                  "2012",
                                  2013",
                                  "2014",
                                  "2015",
                                  2016"))

DB1$metaanalysis_included        <- revalue(DB1$metaanalysis_included,
c("No"="No",
                                  "Metanalisis included"="Meta-analysis included"))
levels(DB1$funding_industry)     <- c("No","Funded by Pharma")
levels(DB1$funding_academic)     <- c("No","Funded by Academic")
levels(DB1$metaanalysis_included) <- c("No","Metanalisis included")
```

```
levels(DB1$clinical_trials)       <- c("No","Analysis of clinical_trials")
levels(DB1$observational_studies) <- c("No","Analysis of observational studies")
levels(DB1$systematic_reviews)    <- c("No","Analysis of systematic reviews")
levels(DB1$economic_studies)      <- c("No","Analysis of economic studies")

DB1$topic_treatment <- mapvalues(DB1$topic,
from = c("Comorbidities",
                        "Economic analysis",
                        "Pathogeny",
                        "Treatment"),
                        to = c("0", "0", "0", "Treatment"))
DB1$topic_Pathogeny <- mapvalues(DB1$topic,
from = c("Comorbidities",
                        "Economic analysis",
                        "Pathogeny",
                        "Treatment"),
                        to = c("0", "0", "Pathogeny", "0"))
DB1$topic_Economic_analysis<- mapvalues(DB1$topic,
from = c("Comorbidities",
                        "Economic analysis",
                        "Pathogeny",
                        "Treatment"), to = c("0", "Economic analysis", "0", "0"))
DB1$topic_Comorbidities <- mapvalues(DB1$topic,
from = c("Comorbidities",
                        "Economic analysis",
                        "Pathogeny",
                        "Treatment"), to = c("Comorbidities", "0", "0", "0"))

file2 <- read.csv2("authors_metadata_SCOPUS.csv",
           sep = ";",
           dec= ".",
           stringsAsFactors = TRUE,
            header = TRUE)
DB2 <- as.data.frame(file2)

DB_merged <- merge(DB1, DB2, by.x="article_Id", by.y="article_Id")

############################################################
######## 1: Multinomial logistic regression    ########
############################################################

##########  ordinal regression dataset ---------

DB1_ordinal_regression_dataset      <- na.omit(DB1[,c("AMSTAR_levels_2",
```

```
                                                "page_count",
                                                "conflict_of_interest",
                                                "X5.Year.Impact.Factor",
                                                "funding_academic",
                                                "metaanalysis_included",
                                                "Article.Influence.Score")])

DB_merge_ordinal_regression_dataset <- na.omit(DB_merged[,c("AMSTAR_levels_2",
                                                "page_count",
                                                "conflict_of_interest",
                                                "X5.Year.Impact.Factor",
                                                "funding_academic",
                                                "metaanalysis_included",
                                                "Article.Influence.Score",
                                                "Documents",
                                                "Citations",
                                                "Co.authors",
                                                "H.index_author")])

##########  ordinal regression model ---------

model_ordinal_regression            <- orm(AMSTAR_levels_2 ~ log(page_count,2)+
                                          conflict_of_interest+
                                          X5.Year.Impact.Factor+
                                          funding_academic+
                                          metaanalysis_included,
                                          Article.Influence.Score,
                                          data = DB1_ordinal_regression_dataset)

########## ordinal regression results OR, 95\%CI ---------

(ci <- confint(model_ordinal_regression))
exp(coef(model_ordinal_regression))
exp(cbind(OR = coef(model_ordinal_regression), ci))

#############################################################################
######## 2: model Validation. Repeated k-fold Cross Validation.  ########
#############################################################################

# load the library
library(caret)
library(klaR)
library(e1071)
```

```
# define training control
train_control<- trainControl(method="repeatedcv", number=10, repeats=3)

# train the model
model<- train(AMSTAR_levels_2~page_count+
conflict_of_interest+
            funding_academic+
            metaanalysis_included,
            data=DB1_ordinal_regression_dataset,
            trControl=train_control,
            method="nb")

# summarize results
print(model)


#############################################################
######## 3: ordinal regression probability plots   ########
#############################################################

library(nnet)
library(memisc)

##########   ---------

# multiplot function
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  library(grid)

  # Make a list from the ... arguments and plotlist
  plots <- c(list(...), plotlist)

  numPlots = length(plots)

  # If layout is NULL, then use 'cols' to determine layout
  if (is.null(layout)) {
    # Make the panel
    # ncol: Number of columns of plots
    # nrow: Number of rows needed, calculated from # of cols
    layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                     ncol = cols, nrow = ceiling(numPlots/cols))
  }

 if (numPlots==1) {
```

```
    print(plots[[1]])

  } else {
    # Set up the page
    grid.newpage()
    pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

    # Make each plot, in the correct location
    for (i in 1:numPlots) {
      # Get the i,j matrix positions of the regions that contain this subplot
      matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

      print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                      layout.pos.col = matchidx$col))
    }
   }
  }

###### META-ANALYSES INCLUDED
#### vs X5.Year.Impact.Factor
DB1$AMSTAR_levels_2_new <- relevel(DB1$AMSTAR_levels_2, ref = "high_quality")

#################################################################
##1____##log(page_count,2) x metaanalyses_included   ########
#################################################################
DB1$page_count_2 <- log(DB1$page_count,2)
test<- multinom(AMSTAR_levels_2_new ~ page_count_2 +
           metaanalysis_included, data = DB1)
summary(test)
z<- summary(test)$coefficients/summary(test)$standard.errors
#2-tailed z test
p<- (1-pnorm(abs(z), 0, 1))*2
## extract the coefficients from the model and exponentiate
exp(coef(test))
head(pp <- fitted(test))

#Another way to understand the model using the predicted
# probabilities is to look at the
# averaged predicted probabilities for different values of the
# continuous predictor variable conflict_of_interest within each level
# of funding_industry

dpage_count <- data.frame(metaanalysis_included = rep(c("No", "Meta-analysis included"),
each = 15),
```

```
page_count_2 = rep(c(1:10), 3))

## store the predicted probabilities for each value of ses and write
pp.dpage_count <- cbind(dpage_count,
predict(test, newdata = dpage_count, type = "probs", se = TRUE))

## calculate the mean probabilities within each level of ses
by(pp.dpage_count[, c(3,5,4)], pp.dpage_count$metaanalysis_included, colMeans)

##Sometimes, a couple of plots can convey a good deal amount of information.
#Using the predictions we generated for the pp.Journal.Impact.Factor object above,
#we can plot the predicted probabilities against the AMSTAR_levels_2_new score by the
level of funding_industry
#for different levels of the outcome variable.
## melt data set to long for ggplot2
lpp_1  <- melt(pp.dpage_count,
id.vars = c("metaanalysis_included", "page_count_2"),
            value.name = "probability")
head(lpp_1) # view first few

#####################################################################
##2____##conflict_of_interest x metaanalyses_included   ########
#####################################################################
DB1$AMSTAR_levels_2_new <- relevel(DB1$AMSTAR_levels_2, ref = "high_quality")
test<- multinom(AMSTAR_levels_2_new ~ conflict_of_interest +
           metaanalysis_included, data = DB1)
summary(test)

## 2-tailed z test
z <- summary(test)$coefficients/summary(test)$standard.errors
p <- (1-pnorm(abs(z), 0, 1))*2

## extract the coefficients from the model and exponentiate
exp(coef(test))
head(pp <- fitted(test))

#Another way to understand the model using the predicted
# probabilities is to look at the
# averaged predicted probabilities for different values of
# the continuous predictor variable conflict_of_interest within
# each level of funding_industry

dconflict_of_interest_2 <- data.frame(metaanalysis_included = rep(c("No", "Meta-analysis
included"), each = 24),
```

```
conflict_of_interest = rep(c(0:15), 3))

## store the predicted probabilities for each value of ses and write
pp.dconflict_of_interest_2 <- cbind(dconflict_of_interest_2,
predict(test, newdata = dconflict_of_interest_2, type = "probs", se = TRUE))

## calculate the mean probabilities within each level of ses
by(pp.dconflict_of_interest_2[, c(3,5,4)],
pp.dconflict_of_interest_2$funding_academic, colMeans)

## Sometimes, a couple of plots can convey a good deal amount of information.
# Using the predictions we generated for the pp.Journal.Impact.Factor object above,
# we can plot the predicted probabilities against the
# AMSTAR_levels_2_new score by the level of funding_industry
# for different levels of the outcome variable.
## melt data set to long for ggplot2

lpp_2  <- melt(pp.dconflict_of_interest_2,
 id.vars = c("metaanalysis_included", "conflict_of_interest"), value.name = "probability")
head(lpp_2) # view first few


#########################################################################
##3____##Article.Influence.Score x metaanalyses_included   ########
#########################################################################
DB1$AMSTAR_levels_2_new <- relevel(DB1$AMSTAR_levels_2, ref = "high_quality")
test<- multinom(AMSTAR_levels_2_new ~ Article.Influence.Score +
         metaanalysis_included, data = DB1)
summary(test)

## 2-tailed z test
z <- summary(test)$coefficients/summary(test)$standard.errors
p <- (1-pnorm(abs(z), 0, 1))*2

## extract the coefficients from the model and exponentiate
exp(coef(test))
head(pp <- fitted(test))

#Another way to understand the model using the predicted probabilities
# is to look at the averaged predicted probabilities for different values
# of the continuous predictor variable conflict_of_interest within
# each level of funding_industry
```

```
dArticle.Influence.Score <- data.frame(metaanalysis_included = rep(c("No", "Meta-analysis
included"), each = 24),
Article.Influence.Score = rep(c(0:15), 3))

## store the predicted probabilities for each value of ses and write
pp.dArticle.Influence.Score <- cbind(dArticle.Influence.Score,
predict(test, newdata = dArticle.Influence.Score, type = "probs", se = TRUE))

## calculate the mean probabilities within each level of ses
by(pp.dArticle.Influence.Score[, c(3,5,4)],
pp.dArticle.Influence.Score$metaanalysis_included, colMeans)

##Sometimes, a couple of plots can convey a good deal amount of information.
#Using the predictions we generated for the pp.Journal.Impact.Factor object above,
#we can plot the predicted probabilities against the AMSTAR_levels_2_new score by the
level of funding_industry
#for different levels of the outcome variable.
## melt data set to long for ggplot2
lpp_3                                   <- melt(pp.dArticle.Influence.Score, id.vars
= c("metaanalysis_included", "Article.Influence.Score"), value.name = "probability")
head(lpp_3) # view first few


#####################################################################
##4____##X5.Year.Impact.Factor x metaanalyses_included   ########
#####################################################################
test                                    <- multinom(AMSTAR_levels_2_new ~ X5.Year.Impact.Factor
+ metaanalysis_included, data = DB1)
summary(test)

## 2-tailed z test
z                                       <- summary(test)$coefficients/summary(test)$standard.error
p                                       <- (1-pnorm(abs(z), 0, 1))*2

## extract the coefficients from the model and exponentiate
exp(coef(test))
head(pp <- fitted(test))

#Another way to understand the model using the predicted probabilities is to look at
the
#averaged predicted probabilities for different values of the continuous predictor variable
conflict_of_interest within each level of funding_industry
```

```
dX5.Year.Impact.Factor                    <- data.frame(metaanalysis_included = rep(c("No",
"Meta-analysis included"), each = 45),X5.Year.Impact.Factor = rep(c(0.5:30), 3))

## store the predicted probabilities for each value of ses and write
pp.X5.Year.Impact.Factor                  <- cbind(dX5.Year.Impact.Factor, predict(test,
newdata = dX5.Year.Impact.Factor, type = "probs", se = TRUE))

## calculate the mean probabilities within each level of ses
by(pp.X5.Year.Impact.Factor[, c(3,5,4)], pp.X5.Year.Impact.Factor$metaanalysis_included,
colMeans)

##Sometimes, a couple of plots can convey a good deal amount of information.
#Using the predictions we generated for the pp.Journal.Impact.Factor object above,
#we can plot the predicted probabilities against the AMSTAR_levels_2_new score by the
level of funding_industry
#for different levels of the outcome variable.
## melt data set to long for ggplot2
lpp_4                                     <- melt(pp.X5.Year.Impact.Factor, id.vars =
c("metaanalysis_included", "X5.Year.Impact.Factor"), value.name = "probability")
head(lpp_4) # view first few


############################################################
##5____##log(page_count,2) x funding_academic   ########
############################################################
DB1$page_count_2                          <- log(DB1$page_count,2)
test                                      <- multinom(AMSTAR_levels_2_new ~ page_count_2
+ funding_academic, data = DB1)
summary(test)

## 2-tailed z test
z                                         <- summary(test)$coefficients/summary(test)$standard.er
p                                         <- (1-pnorm(abs(z), 0, 1))*2

## extract the coefficients from the model and exponentiate
exp(coef(test))
head(pp <- fitted(test))

#Another way to understand the model using the predicted probabilities is to look at
the
#averaged predicted probabilities for different values of the continuous predictor variable
conflict_of_interest within each level of funding_industry
dpage_count_2                             <- data.frame(funding_academic = rep(c("No",
"Funded by Academic"), each = 15),page_count_2 = rep(c(1:10), 3))
```

```
## store the predicted probabilities for each value of ses and write
pp.dpage_count_2                          <- cbind(dpage_count_2, predict(test, newdata
= dpage_count_2, type = "probs", se = TRUE))

## calculate the mean probabilities within each level of ses
by(pp.dpage_count_2[, c(3,5,4)], pp.dpage_count_2$funding_academic, colMeans)

##Sometimes, a couple of plots can convey a good deal amount of information.
#Using the predictions we generated for the pp.Journal.Impact.Factor object above,
#we can plot the predicted probabilities against the AMSTAR_levels_2_new score by the
level of funding_industry
#for different levels of the outcome variable.
## melt data set to long for ggplot2
lpp_5                                     <- melt(pp.dpage_count_2, id.vars = c("funding_academic",
"page_count_2"), value.name = "probability")
head(lpp_5) # view first few

############################################################
##6____##conflict_of_interest x funding_academic   ########
############################################################
DB1$AMSTAR_levels_2_new                   <- relevel(DB1$AMSTAR_levels_2, ref = "high_quality")
test                                      <- multinom(AMSTAR_levels_2_new ~ conflict_of_interest
+ funding_academic, data = DB1)
summary(test)

## 2-tailed z test
z                                         <- summary(test)$coefficients/summary(test)$standard.error
p                                         <- (1-pnorm(abs(z), 0, 1))*2
## extract the coefficients from the model and exponentiate
exp(coef(test))
head(pp <- fitted(test))

#Another way to understand the model using the predicted probabilities is to look at
the
#averaged predicted probabilities for different values of the continuous predictor variable
conflict_of_interest within each level of funding_industry
dconflict_of_interest_new                 <- data.frame(funding_academic = rep(c("No",
"Funded by Academic"), each = 24),conflict_of_interest = rep(c(0:15), 3))

## store the predicted probabilities for each value of ses and write
pp.dconflict_of_interest_new              <- cbind(dconflict_of_interest_new, predict(test,
newdata = dconflict_of_interest_new, type = "probs", se = TRUE))
```

```
## calculate the mean probabilities within each level of ses
by(pp.dconflict_of_interest_new[, c(3,5,4)], pp.dconflict_of_interest_new$funding_academic,
colMeans)

##Sometimes, a couple of plots can convey a good deal amount of information.
#Using the predictions we generated for the pp.Journal.Impact.Factor object above,
#we can plot the predicted probabilities against the AMSTAR_levels_2_new score by the
level of funding_industry
#for different levels of the outcome variable.
## melt data set to long for ggplot2
lpp_6                                   <- melt(pp.dconflict_of_interest_new, id.vars
= c("funding_academic", "conflict_of_interest"), value.name = "probability")
head(lpp_6) # view first few


#################################################################
##7____##Article.Influence.Score x funding_academic   ########
#################################################################
DB1$AMSTAR_levels_2_new                 <- relevel(DB1$AMSTAR_levels_2, ref = "high_quality")
test                                    <- multinom(AMSTAR_levels_2_new ~ Article.Influence.Sco
+ funding_academic, data = DB1)
summary(test)

## 2-tailed z test
z                                       <- summary(test)$coefficients/summary(test)$standard.er
p                                       <- (1-pnorm(abs(z), 0, 1))*2

## extract the coefficients from the model and exponentiate
exp(coef(test))
head(pp <- fitted(test))

#Another way to understand the model using the predicted probabilities is to look at
the
#averaged predicted probabilities for different values of the continuous predictor variable
conflict_of_interest within each level of funding_industry
dArticle.Influence.Score_2              <- data.frame(funding_academic = rep(c("No",
"Funded by Academic"), each = 24),Article.Influence.Score = rep(c(0:15), 3))

## store the predicted probabilities for each value of ses and write
pp.dArticle.Influence.Score_2           <- cbind(dArticle.Influence.Score_2, predict(test,
newdata = dArticle.Influence.Score_2, type = "probs", se = TRUE))

## calculate the mean probabilities within each level of ses
```

```
by(pp.dArticle.Influence.Score_2[, c(3,5,4)], pp.dArticle.Influence.Score_2$funding_academic,
colMeans)

##Sometimes, a couple of plots can convey a good deal amount of information.
#Using the predictions we generated for the pp.Journal.Impact.Factor object above,
#we can plot the predicted probabilities against the AMSTAR_levels_2_new score by the
level of funding_industry
#for different levels of the outcome variable.
## melt data set to long for ggplot2
lpp_7                                   <- melt(pp.dArticle.Influence.Score_2, id.vars
= c("funding_academic", "Article.Influence.Score"), value.name = "probability")
head(lpp_7) # view first few


############################################################
##8____##X5.Year.Impact.Factor x funding_academic   ########
############################################################
DB1$AMSTAR_levels_2_new                 <- relevel(DB1$AMSTAR_levels_2, ref = "high_quality")
test                                    <- multinom(AMSTAR_levels_2_new ~ X5.Year.Impact.Factor
+ funding_academic, data = DB1)
summary(test)

## 2-tailed z test
z                                       <- summary(test)$coefficients/summary(test)$standard.error
p                                       <- (1-pnorm(abs(z), 0, 1))*2

## extract the coefficients from the model and exponentiate
exp(coef(test))
head(pp <- fitted(test))

#Another way to understand the model using the predicted probabilities is to look at
the
#averaged predicted probabilities for different values of the continuous predictor variable
conflict_of_interest within each level of funding_industry
dX5.Year.Impact.Factor_2                <- data.frame(funding_academic = rep(c("No",
"Funded by Academic"), each = 45),X5.Year.Impact.Factor = rep(c(0.5:30), 3))

## store the predicted probabilities for each value of ses and write
pp.dX5.Year.Impact.Factor_2             <- cbind(dX5.Year.Impact.Factor_2, predict(test,
newdata = dX5.Year.Impact.Factor_2, type = "probs", se = TRUE))

## calculate the mean probabilities within each level of ses
by(pp.dX5.Year.Impact.Factor_2[, c(3,5,4)], pp.dX5.Year.Impact.Factor_2$funding_academic,
colMeans)
```

```
##Sometimes, a couple of plots can convey a good deal amount of information.
#Using the predictions we generated for the pp.Journal.Impact.Factor object above,
#we can plot the predicted probabilities against the AMSTAR_levels_2_new score by the
level of funding_industry
#for different levels of the outcome variable.
## melt data set to long for ggplot2
lpp_8                                   <- melt(pp.dX5.Year.Impact.Factor_2, id.vars
= c("funding_academic", "X5.Year.Impact.Factor"), value.name = "probability")
head(lpp_8) # view first few


lpp_1$variable                          <- factor(lpp_1$variable, levels=c("high_quality",
"moderate_quality","low_quality"))
lpp_1$variable                          <- relabel(lpp_1$variable, "high_quality"="High","modera

lpp_2$variable                          <- factor(lpp_2$variable, levels=c("high_quality",
"moderate_quality","low_quality"))
lpp_2$variable                          <- relabel(lpp_2$variable, "high_quality"="High","modera

lpp_3$variable                          <- factor(lpp_3$variable, levels=c("high_quality",
"moderate_quality","low_quality"))
lpp_3$variable                          <- relabel(lpp_3$variable, "high_quality"="High","modera

lpp_4$variable                          <- factor(lpp_4$variable, levels=c("high_quality",
"moderate_quality","low_quality"))
lpp_4$variable                          <- relabel(lpp_4$variable, "high_quality"="High","modera

lpp_5$variable                          <- factor(lpp_5$variable, levels=c("high_quality",
"moderate_quality","low_quality"))
lpp_5$variable                          <- relabel(lpp_5$variable, "high_quality"="High","modera

lpp_6$variable                          <- factor(lpp_6$variable, levels=c("high_quality",
"moderate_quality","low_quality"))
lpp_6$variable                          <- relabel(lpp_6$variable, "high_quality"="High","modera

lpp_7$variable                          <- factor(lpp_7$variable, levels=c("high_quality",
"moderate_quality","low_quality"))
lpp_7$variable                          <- relabel(lpp_7$variable, "high_quality"="High","modera

lpp_8$variable                          <- factor(lpp_8$variable, levels=c("high_quality",
"moderate_quality","low_quality"))
lpp_8$variable                          <- relabel(lpp_8$variable, "high_quality"="High","modera
```

```
p1                                          <- ggplot(lpp_1,aes(x = page_count_2, y = value,
colour = metaanalysis_included)) +
                                                    geom_line(aes(linetype = metaanalysis_included,
color = metaanalysis_included)) +
                                                    scale_linetype_manual(values = c("solid",
"dotted")) +
                                                    scale_color_manual(values = c('#E69F00',
'#E69F00')) +
                                                    scale_size_manual(values =c(3, 5)) +
                                                    facet_grid(variable ~ ., scales = "fixed")
+
                                                    theme(legend.position = "top") + theme(legend.title
= element_blank()) +
                                                    theme(panel.grid.minor = element_line(colour
= "grey", linetype = "dotted")) +
                                                    theme(panel.grid.major = element_line(colour
= "grey")) +
                                                    labs(x = "log2(page count)") +
                                                    scale_y_continuous(limits = c(0, 1),
breaks = seq(0, 1, by = 0.20))

p2                                          <- ggplot(lpp_2,aes(x = conflict_of_interest,
y = value, colour = metaanalysis_included)) +
                                                    geom_line(aes(linetype = metaanalysis_included,
color = metaanalysis_included)) +
                                                    scale_linetype_manual(values = c("solid",
"dotted")) + scale_color_manual(values =c('#458B74', '#458B74')) +
                                                    scale_size_manual(values = c(3, 5)) +
                                                    facet_grid(variable ~ ., scales ="fixed")
+
                                                    theme(legend.position = "top") + theme(legend.title
= element_blank()) +
                                                    theme(panel.grid.minor = element_line(colour
= "grey", linetype = "dotted")) +
                                                    theme(panel.grid.major = element_line(colour
= "grey")) +
                                                    labs(x = "Authors with conflict of interest")
+
                                                    scale_y_continuous(limits = c(0, 1),
breaks = seq(0, 1, by = 0.20))

p3                                           <-ggplot(lpp_3, aes(x = Article.Influence.Score,
y = value, colour = metaanalysis_included)) +
```

```
                                                    geom_line(aes(linetype=metaanalysis_included,
color=metaanalysis_included))+
                                                    scale_linetype_manual(values=c("solid",
"dotted"))+
                                                    scale_color_manual(values=c('#6495ED','#6495ED'))+
                                                    scale_size_manual(values=c(3, 5))+
                                                    facet_grid(variable ~ ., scales="fixed")+
                                                    theme(legend.position="top")+
                                                    theme(legend.title=element_blank())+
                                                    theme(panel.grid.minor = element_line(colour
= "grey", linetype = "dotted"))+
                                                    theme(panel.grid.major = element_line(colour
= "grey"))+
                                                    labs(x="Article Influence Score")+scale_y_continuo
= seq(0, 1, by = 0.20))

p4                                          <-ggplot(lpp_4, aes(x = X5.Year.Impact.Factor,
y = value, colour = metaanalysis_included)) +
                                                    geom_line(aes(linetype=metaanalysis_included,
color=metaanalysis_included))+
                                                    scale_linetype_manual(values=c("solid",
"dotted"))+
                                                    scale_color_manual(values=c('#A2CD5A','#A2CD5A'))+
                                                    scale_size_manual(values=c(3, 5))+
                                                    facet_grid(variable ~ ., scales="fixed")+
                                                    theme(legend.position="top")+
                                                    theme(legend.title=element_blank())+
                                                    theme(panel.grid.minor = element_line(colour
= "grey", linetype = "dotted"))+
                                                    theme(panel.grid.major = element_line(colour
= "grey"))+
                                                    labs(x="5 Year Impact Factor")+
                                                    scale_y_continuous(limits=c(0,1),breaks
= seq(0, 1, by = 0.20))

p5                                          <-ggplot(lpp_5,
                                                    aes(x = page_count_2, y = value, colour
= funding_academic)) +
                                                    geom_line(aes(linetype=funding_academic,
color=funding_academic))+
                                                    scale_linetype_manual(values=c("solid",
"dotted"))+
                                                    scale_color_manual(values=c('#BB0000','#BB0000'))+
                                                    scale_size_manual(values=c(3, 5))+
```

```
                                        facet_grid(variable ~ .,scales="fixed")+
                                        theme(legend.position="top")+
                                        theme(legend.title=element_blank())+
                                        theme(panel.grid.minor = element_line(colour
= "grey", linetype = "dotted"))+
                                        theme(panel.grid.major = element_line(colour
= "grey"))+
                                        labs(x="log2(page count)")+
                                        scale_y_continuous(limits=c(0,1),breaks
= seq(0, 1, by = 0.20))

p6                                  <-ggplot(lpp_6, aes(x = conflict_of_interest,
y = value, colour = funding_academic)) +
                                        geom_line(aes(linetype=funding_academic,
color=funding_academic))+
                                        scale_linetype_manual(values=c("solid",
"dotted"))+
                                        scale_color_manual(values=c('#006400','#006400'))+
                                        scale_size_manual(values=c(3, 5))+
                                        facet_grid(variable ~ ., scales="fixed")+
                                        theme(legend.position="top")+
                                        theme(legend.title=element_blank())+
                                        theme(panel.grid.minor = element_line(colour
= "grey", linetype = "dotted"))+
                                        theme(panel.grid.major = element_line(colour
= "grey"))+
                                        labs(x="Authors with conflict of interest")+
                                        scale_y_continuous(limits=c(0,1),breaks
= seq(0, 1, by = 0.20))

p7                                  <-ggplot(lpp_7, aes(x = Article.Influence.Score,
y = value, colour = funding_academic)) +
                                        geom_line(aes(linetype=funding_academic,
color=funding_academic))+
                                        scale_linetype_manual(values=c("solid",
"dotted"))+
                                        scale_color_manual(values=c('#00009B','#00009B'))+
                                        scale_size_manual(values=c(3, 5))+
                                        facet_grid(variable ~ ., scales="fixed")+
                                        theme(legend.position="top")+
                                        theme(legend.title=element_blank())+
                                        theme(panel.grid.minor = element_line(colour
= "grey", linetype = "dotted"))+
```

```
                                                    theme(panel.grid.major = element_line(colour
= "grey"))+
                                                    labs(x="Article Influence Score")+
                                                    scale_y_continuous(limits=c(0,1),breaks
= seq(0, 1, by = 0.20))

p8                                              <-ggplot(lpp_8, aes(x = X5.Year.Impact.Factor,
y = value, colour = funding_academic)) +
                                                    geom_line(aes(linetype=funding_academic,
color=funding_academic))+
                                                    scale_linetype_manual(values=c("solid",
"dotted"))+
                                                    scale_color_manual(values=c('#006400','#006400'))+
                                                    scale_size_manual(values=c(3, 5))+
                                                    facet_grid(variable ~ ., scales="fixed")+
                                                    theme(legend.position="top")+
                                                    theme(legend.title=element_blank())+
                                                    theme(panel.grid.minor = element_line(colour
= "grey", linetype = "dotted"))+
                                                    theme(panel.grid.major = element_line(colour
= "grey"))+
                                                    labs(x="5 Year Impact Factor")+scale_y_continuous
= seq(0, 1, by = 0.20))


multiplot(p1,p5,p2,p6,p3,p7,p4,p8,cols=4)

#' ---
#' title: "Methodological appraisal of systematic reviews and meta-analyses on psoriasis:
role of funding sources, conflict of interest and pressure to publish"
#' author: "Juan Ruano"
#' date: "30 Sep 2016"
#' institutions: Department of Dermatology, IMIBIC/Reina Sofia University Hospital/University
of Cordoba, Cordoba, Spain
#' analyse: 01_Bubble plot pharma funding  vs  AMSTAR score
#' ---
#'
#' R version 3.3.1 (2016-06-21)
#' Platform: x86_64-apple-darwin13.4.0 (64-bit)
#' Running under: OS X 10.9.5 (Mavericks)


########  R packages ----------------
```

```
library(plyr)
library(ggplot2)
library(cowplot)

########  environment setting ----------------
# For RStudio only

setwd(dirname(rstudioapi::callFun("getActiveDocumentContext")$path))


######## Read .csv files -------------------------------

####  main DB -----

file1<-read.csv2("AMSTARfinal_metadatos_articulos_con_topics.csv",
                 sep = ";",
                 dec= ".",
                 stringsAsFactors = TRUE,
                 header = TRUE)
DB1<-as.data.frame(file1)

####  scopus data -----

file5<-read.csv2("authors_metadata_SCOPUS.csv",
                 sep = ";",
                 dec= ".",
                 stringsAsFactors = TRUE,
                 header = TRUE)
DB_scopus<-as.data.frame(file5)

########  Tiding dataset --------------------------------

DB1$Total.Cites                   <- as.numeric(DB1$Total.Cites)
DB1$Cited.Half.life               <- as.numeric(DB1$Cited.Half.life)
DB1$Citing.Half.life              <- as.numeric(DB1$Citing.Half.life)
DB1$Normalized.Eigenfactor        <- as.numeric(DB1$Normalized.Eigenfactor)
DB1$num_institutions              <- as.numeric(DB1$num_institutions)
DB1$conflict_of_interest          <- as.numeric(DB1$conflict_of_interest)

DB1$funding_academic              <- as.factor(DB1$funding_academic)
DB1$funding_industry              <- as.factor(DB1$funding_industry)
DB1$funding                       <- as.factor(DB1$funding)

DB1$metaanalysis_included         <- factor(DB1$metaanalysis_included)
```

```
DB1$AMSTAR_levels                 <- factor(DB1$AMSTAR_levels, levels=c("low_quality",
"moderate_quality", "high_quality"))
DB1$AMSTAR_levels_2               <- factor(DB1$AMSTAR_levels_2, levels=c("low_quality",
"moderate_quality", "high_quality"))
DB1$SJR.Best.Quartile             <- factor(DB1$SJR.Best.Quartile, levels=c("Q1", "Q2",
"Q3", "Q4"))
DB1$Country                       <- factor(DB1$Country, levels=c("Brazil", "Canada",
"China", "Germany", "India", "Ireland", "Netherlands", "Sweden", "Switzerland", "United
Kingdom"))
DB1$topic                         <- factor(DB1$topic, levels=c("Comorbidities", "Economic
analyses", "Pathogeny", "Treatment"))
DB1$year                          <- factor(DB1$year, levels=c("1999", "2000", "2001",
"2002", "2003", "2004", "2005", "2006", "2007", "2008", "2009", "2010", "2011", "2012",
"2013", "2014", "2015", "2016"))

levels(DB1$funding_industry)      <- c("No","Funded by Pharma")
levels(DB1$funding_academic)      <- c("No","Funded by Academic")
levels(DB1$metaanalysis_included) <- c("No","Metanalisis included")
levels(DB1$clinical_trials)       <- c("No","Analyses of clinical_trials")
levels(DB1$observational_studies) <- c("No","Analyses of observational studies")
levels(DB1$systematic_reviews)    <- c("No","Analyses of systematic reviews")
levels(DB1$economic_studies)      <- c("No","Analyses of economic studies")
levels(DB1$AMSTAR_levels_2)       <- c("low quality", "moderate quality", "high quality")
levels(DB1$AMSTAR_levels_2)       <- factor(DB1$AMSTAR_levels_2, levels=c("low quality",
"moderate quality", "high quality"))



DB1$topic_treatment               <- mapvalues(DB1$topic, from = c("Comorbidities", "Economic
analyses", "Pathogeny", "Treatment"), to = c("0", "0", "0", "Treatment"))
DB1$topic_Pathogeny               <- mapvalues(DB1$topic, from = c("Comorbidities", "Economic
analyses", "Pathogeny", "Treatment"), to = c("0", "0", "Pathogeny", "0"))
DB1$topic_Economic_analysis       <- mapvalues(DB1$topic, from = c("Comorbidities", "Economic
analyses", "Pathogeny", "Treatment"), to = c("0", "Economic analyses", "0", "0"))
DB1$topic_Comorbidities           <- mapvalues(DB1$topic, from = c("Comorbidities", "Economic
analyses", "Pathogeny", "Treatment"), to = c("Comorbidities", "0", "0", "0"))



DB_bubble<-merge(DB1, DB_scopus, by.x="article_Id", by.y="article_Id")


##########  bubble plot ----------------
```

```
DB_bubble$pharma_name<-factor(DB_bubble$pharma_name, levels=c("Galderma", "Pfizer", "Novartis",
"AbbVie", "LEO Pharma", "Boots Healthcare International", "Janssen-Cilag", "Wyeth Pharma",
"UCB Pharma", "MSD"))
ggplot(na.omit(DB_bubble[,c("AMSTAR_consensus_2","pharma_name")]), aes(x = AMSTAR_consensus_2,
y =pharma_name)) +
  geom_count(colour="royalblue4")+
  scale_x_continuous(breaks=c(1,2,3,4,5,6,7,8,9,10,11))+
  theme_minimal()+
  theme(legend.position="none")+
  theme(text = element_text(color = "gray10"),panel.grid.major = element_line(color =
"gray80", size = 0.5),panel.grid.major.x = element_blank())+
  xlab("AMSTAR score")+
  ylab("Pharmaceutical Companies")+
  geom_vline(xintercept=c(4,8), linetype="dotted", colour = "blue", size = 1, alpha =
.4)+
  scale_size(range = c(5,30))

##########  funding pharma rank ----------------

summary(DB_bubble$pharma_name)

##########  top institutions ranked by number of papers and AMSTAR score ----------------

count_authors<-unique(count(DB_bubble$Institution))
write.table(count_authors, file="count_authors.xls")


median_per_institutio<-unique(ddply(DB_bubble, .(Institution), summarize,  Rate1=median(AMSTAR_conse
write.table(median_per_institutio, file="median_per_institutio.xls")


#' ---
#' title: "Methodological appraisal of systematic reviews and meta-analyses on psoriasis:
role of funding sources, conflict of interest and pressure to publish"
#' author: "Juan Ruano"
#' date: "30 Sep 2016"
#' institutions: Department of Dermatology, IMIBIC/Reina Sofia University Hospital/University
of Cordoba, Cordoba, Spain
#' analyse: 08_Polar plots of AMSTAR scores by journal (ranked by mean values)
#' ---

##### world map plot AMSTAR scores ----------------
##  R packages --------
library(utils)
```

```
library(base)
library(reshape2)
library(stats)
library(ggplot2)

##  R functions --------
# Using the new ggproto mechanism available in ggplot2 2.0.0, coord_radar can be defined
as

coord_radar <- function (theta = "x", start = 0, direction = 1) {
    theta <- match.arg(theta, c("x", "y"))
        r <- if (theta == "x")
          "y"
        else
          "x"
    ggproto("CordRadar", CoordPolar, theta = theta, r = r, start = start,
        direction = sign(direction),
        is_linear = function(coord) FALSE)
  }


##  file environment setting --------
# For RStudio only

setwd(dirname(rstudioapi::callFun("getActiveDocumentContext")$path))


##  file read --------

file2               <- utils::read.csv2("dbPolarPlots.csv", sep=";", dec=".", header=TRUE)
DB_polar            <- base::as.data.frame(file2)
DB_polar_treatment  <- base::subset(DB_polar, topic=="treatment")


##  tyding dataset ONLY SR and MA for intervention studies --------

DB_polar_treatment$Journal_2 <- base::factor(DB_polar_treatment$Journal_2, levels = c("J
Invest Dermatol","Clin Exp Dermatol", "Int J Dermatol", "JAMA Dermatol (Arch Dermatol)","J
Am Acad Dermatol","Br J Dermatol","Dermatology","Arch Dermatol Res", "J Dermatol","J
Cutan Med Surg","Acta Derm Venereol","Photodermatol Photoimmunol Photomed","J Eur Acad
Dermatol Venereol","J Dtsch Dermatol Ges","J Dermatolog Treat","Pediatr Dermatol","Indian
J Dermatol Venereol Leprol","Dermatol Online J","Dermatol Ther (Heidelb)","Am J Clin
Dermatol"))
```

```
DB_polar_treatment_melted    <- reshape2::melt(DB_polar_treatment[c(2:13,15,16)],  id=c("AMSTAR_leve
"Journal_2", "num_publications"))
DB_polar_treatment_aggregate <- stats::aggregate(value ~ Journal_2 + variable+num_publications,
data = DB_polar_treatment_melted, sum)


## polar plot --------

ggplot(DB_polar_treatment_aggregate, aes(x = variable, y = value/num_publications)) +
  geom_polygon(aes(group = Journal_2,color = Journal_2, size = num_publications),  fill
= NA) +
  geom_line(aes(group = Journal_2, color = Journal_2, size = num_publications))+
  facet_wrap(~ Journal_2) +
  theme(axis.ticks.x = element_blank(),
        axis.text.x = element_text(size = rel(0.6)),
        axis.ticks.y = element_blank(),
        axis.text.y = element_blank())+
  xlab("") + ylab("") +
  guides(color = "none") +
  labs(size ="number of publications")+
  coord_radar()
```

R code for paper #3

```
#################################################################################################
# ---
# title: "Most systematic reviews of high methodological quality on psoriasis interventions
are classified as high risk of bias using ROBIS tool"
# author: "Juan Ruano"
# date: "16 May 2017"
# institutions: Department of Dermatology, IMIBIC/Reina Sofia University Hospital/University
of Cordoba, Cordoba, Spain
# analysis: Statistical analyses: PCA, Likert scales, correlation matrix and network,
and Radviz plots
# ---
#
# R version 3.3.1 (2016-06-21)
# Platform: x86_64-apple-darwin13.4.0 (64-bit)
# Running under: OS X 10.9.5 (Mavericks)
#
#################################################################################################

########  read .csv files ----------------

file2 <- read.csv2("ROBIS_vs_AMSTAR_pca.csv",
                   sep = ";",
                   dec= ".",
                   stringsAsFactors = TRUE,
                   header = TRUE)
DB2<-as.data.frame(file2)
names(DB2)
head(DB2)

df_amstar <- DB2[, 5:15]
res.pca <- PCA(df, graph = FALSE, scale=T)
round(res.pca$ind$contrib, 4)

DB2$AMSTAR_levels_2 <- factor(DB2$AMSTAR_levels_2,
                              levels = c("high_quality", "moderate_quality","low_quality"
))


#########################################################################
######## 1: PCA ----------------
######## packages ----------------
```

```
library("ggfortify")
library("cluster")
library("reshape2")
library("FactoMineR")
library("factoextra")

######## PCA amstar ----------------

DB2_PCA_amstar <- na.omit(DB2[c(30:40,41,42, 3)])
names(DB2_PCA_amstar)
amstar.pca <- prcomp(DB2_PCA_amstar[c(1:11)])
fviz_screeplot(amstar.pca, ncp=10)

plot_pca_amstar <- autoplot(prcomp(DB2_PCA_amstar[c(1:11)]),
                            data = DB2_PCA_amstar,
                            colour = 'AMSTAR_levels_2',
                            shape="RoB_ROBIS",
                            loadings = TRUE,
                            loadings.label = TRUE,
                            frame = TRUE)
plot_pca_amstar +
  geom_point(aes(shape = factor(RoB_ROBIS),
                 colour = AMSTAR_levels_2)) +
  scale_shape_manual(values=c(24,25))+
  geom_jitter(aes(shape = factor(RoB_ROBIS),
                  colour = AMSTAR_levels_2),
                  position=position_jitter(height=0.009,width = 0.009))

#### plot AMSTAR variable contribution axes

fviz_pca_var(amstar.pca, col.var="contrib") +
  scale_color_gradient2(low="white", mid="blue",high="red", midpoint=10) +
  theme_minimal()

#### plot top10 AMSTAR variables contribution to PC1/PC2

fviz_pca_contrib(amstar.pca,
                 choice = "var",
                 axes = 1:2,
                 fill="coral2",
                 color = "grey") +
  theme(axis.text.x = element_text(size=10, angle=90)) +
  scale_y_continuous(limits = c(0.0, 25.0))
```

```
#### plot top10 reviews contribution to AMSTAR PC1/PC2

fviz_pca_contrib(amstar.pca,
                 choice = "ind",
                 axes = 1:2,
                 top=50,
                 fill="seagreen4",
                 color = "grey") +
  theme(axis.text.x = element_text( size=10, angle=90)) +
  scale_y_continuous(limits = c(0.0, 3.0))

#### plot review contributions to AMSTAR PC1/PC2

fviz_pca_ind(amstar.pca,
             col.ind="contrib",
             jitter = list(what = "label", width = NULL, height = NULL)) +
  scale_color_gradient2(low="white", mid="blue", high="red", midpoint=1) +
  theme_minimal() +
  geom_point(colour="white")


######## PCA robis ----------------

DB2_PCA_robis <- na.omit(DB2[c(4:24,41,42, 3)])

robis.pca <- prcomp(DB2_PCA_robis[c(1:21)])
fviz_screeplot(robis.pca, ncp=10)

plot_pca_robis <- autoplot(prcomp(DB2_PCA_robis[c(1:21)]),
                           data = DB2_PCA_robis,
                           colour = 'RoB_ROBIS',
                           shape="AMSTAR_levels_2",
                           loadings = TRUE,
                           loadings.label = TRUE,
                           frame = TRUE)
plot_pca_robis +
  geom_point(aes(shape = factor(AMSTAR_levels_2),
                 colour = RoB_ROBIS)) +
  scale_shape_manual(values=c(24,4,25))+
  geom_jitter(aes(shape = factor(AMSTAR_levels_2),
                  colour = RoB_ROBIS),
                  position=position_jitter(height=0.009,width = 0.009))
```

```
#### plot ROBIS variable contribution axes

fviz_pca_var(robis.pca, col.var="contrib") +
  scale_color_gradient2(low="white", mid="blue", high="red", midpoint=10) +
  theme_minimal()

#### plot top10 ROBIS variables contribution to PC1/PC2

fviz_pca_contrib(robis.pca, choice = "var", axes = 1:2,  fill="coral2", color = "grey")
+
  theme(axis.text.x = element_text( size=10, angle=90)) +
  scale_y_continuous(limits = c(0.0, 25.0))

#### plot top10 individuals contribution to ROBIS PC1/PC2

fviz_pca_contrib(robis.pca,
                 choice = "ind",
                 axes = 1:2,
                 top=50,fill="seagreen4",
                 color = "grey") +
  theme(axis.text.x = element_text( size=10, angle=90)) +
  scale_y_continuous(limits = c(0.0, 3.0))

#### plot review contributions to ROBIS PC1/PC2
fviz_pca_ind(robis.pca,
             col.ind="contrib",
             jitter = list(what = "label", width = NULL, height = NULL)) +
scale_color_gradient2(low="white", mid="blue", high="red", midpoint=1) +
             theme_minimal() +
             geom_point(colour="white")


################################
#### Plot features that contribute to the classification only ROBIS
robis.pca

df_out_r <- as.data.frame(robis.pca$rotation)
df_out_r$feature <- row.names(df_out_r)

df_out_r

p<-ggplot(df_out_r,aes(x=PC1,y=PC2,label=feature))
p<-p+geom_point(colour="white")+ geom_text(size=3, check_overlap = TRUE)
p
```

```
##### mosaic plot ROBIS vs AMSTAR
mosaicplot(DB2$AMSTAR_levels_2~DB2$RoB_ROBIS,  data=DB2, color = c(23,18), las = 1, xlab="Methodol
quality (AMSTAR)", ylab="Risk of Bias (ROBIS)", main="")



#############################################################################
########## 2: Likert scales ALL ----------------

######## packages --------------------

require(likert)
require(grid)
require(lattice)
require(latticeExtra)
require(HH)
library(plyr)


######## tyding data ----------------

sgbar.likert <- DB2[,c(30:40, 42,3)]
sgbar.likert$Q1_AMSTAR <- as.factor(sgbar.likert$Q1_AMSTAR)
sgbar.likert$Q1_AMSTAR <- revalue(sgbar.likert$Q1_AMSTAR, c("0"="no", "1"="yes"))

sgbar.likert$Q2_AMSTAR <- as.factor(sgbar.likert$Q2_AMSTAR)
sgbar.likert$Q2_AMSTAR <- revalue(sgbar.likert$Q2_AMSTAR, c("0"="no", "1"="yes"))

sgbar.likert$Q3_AMSTAR <- as.factor(sgbar.likert$Q3_AMSTAR)
sgbar.likert$Q3_AMSTAR <- revalue(sgbar.likert$Q3_AMSTAR, c("0"="no", "1"="yes"))

sgbar.likert$Q4_AMSTAR <- as.factor(sgbar.likert$Q4_AMSTAR)
sgbar.likert$Q4_AMSTAR <- revalue(sgbar.likert$Q4_AMSTAR, c("0"="no", "1"="yes"))

sgbar.likert$Q5_AMSTAR <- as.factor(sgbar.likert$Q5_AMSTAR)
sgbar.likert$Q5_AMSTAR <- revalue(sgbar.likert$Q5_AMSTAR, c("0"="no", "1"="yes"))

sgbar.likert$Q6_AMSTAR <- as.factor(sgbar.likert$Q6_AMSTAR)
sgbar.likert$Q6_AMSTAR <- revalue(sgbar.likert$Q6_AMSTAR, c("0"="no", "1"="yes"))

sgbar.likert$Q7_AMSTAR <- as.factor(sgbar.likert$Q7_AMSTAR)
sgbar.likert$Q7_AMSTAR <- revalue(sgbar.likert$Q7_AMSTAR, c("0"="no", "1"="yes"))
```

```
sgbar.likert$Q8_AMSTAR <- as.factor(sgbar.likert$Q8_AMSTAR)
sgbar.likert$Q8_AMSTAR <- revalue(sgbar.likert$Q8_AMSTAR, c("0"="no", "1"="yes"))

sgbar.likert$Q9_AMSTAR <- as.factor(sgbar.likert$Q9_AMSTAR)
sgbar.likert$Q9_AMSTAR <- revalue(sgbar.likert$Q9_AMSTAR, c("0"="no", "1"="yes"))

sgbar.likert$Q10_AMSTAR <- as.factor(sgbar.likert$Q10_AMSTAR)
sgbar.likert$Q10_AMSTAR <- revalue(sgbar.likert$Q10_AMSTAR, c("0"="no", "1"="yes"))

sgbar.likert$Q11_AMSTAR <- as.factor(sgbar.likert$Q11_AMSTAR)
sgbar.likert$Q11_AMSTAR <- revalue(sgbar.likert$Q11_AMSTAR, c("0"="no", "1"="yes"))




sgbar.likert_ROBIS <- DB2[,c(4:24, 42,3)]
sgbar.likert_ROBIS$Q11 <- as.factor(sgbar.likert_ROBIS$Q11)
sgbar.likert_ROBIS$Q11 <- revalue(sgbar.likert_ROBIS$Q11, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q12 <- as.factor(sgbar.likert_ROBIS$Q12)
sgbar.likert_ROBIS$Q12 <- revalue(sgbar.likert_ROBIS$Q12, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q13 <- as.factor(sgbar.likert_ROBIS$Q13)
sgbar.likert_ROBIS$Q13 <- revalue(sgbar.likert_ROBIS$Q13, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q14 <- as.factor(sgbar.likert_ROBIS$Q14)
sgbar.likert_ROBIS$Q14 <- revalue(sgbar.likert_ROBIS$Q14, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q15 <- as.factor(sgbar.likert_ROBIS$Q15)
sgbar.likert_ROBIS$Q15 <- revalue(sgbar.likert_ROBIS$Q15, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q21 <- as.factor(sgbar.likert_ROBIS$Q21)
sgbar.likert_ROBIS$Q21 <- revalue(sgbar.likert_ROBIS$Q21, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q22 <- as.factor(sgbar.likert_ROBIS$Q22)
sgbar.likert_ROBIS$Q22 <- revalue(sgbar.likert_ROBIS$Q22, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))
```

```
sgbar.likert_ROBIS$Q23 <- as.factor(sgbar.likert_ROBIS$Q23)
sgbar.likert_ROBIS$Q23 <- revalue(sgbar.likert_ROBIS$Q23, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q24 <- as.factor(sgbar.likert_ROBIS$Q24)
sgbar.likert_ROBIS$Q24 <- revalue(sgbar.likert_ROBIS$Q24, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q25 <- as.factor(sgbar.likert_ROBIS$Q25)
sgbar.likert_ROBIS$Q25 <- revalue(sgbar.likert_ROBIS$Q25, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q31 <- as.factor(sgbar.likert_ROBIS$Q31)
sgbar.likert_ROBIS$Q31 <- revalue(sgbar.likert_ROBIS$Q31, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q32 <- as.factor(sgbar.likert_ROBIS$Q32)
sgbar.likert_ROBIS$Q32 <- revalue(sgbar.likert_ROBIS$Q32, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q33 <- as.factor(sgbar.likert_ROBIS$Q33)
sgbar.likert_ROBIS$Q33 <- revalue(sgbar.likert_ROBIS$Q33, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q34 <- as.factor(sgbar.likert_ROBIS$Q34)
sgbar.likert_ROBIS$Q34 <- revalue(sgbar.likert_ROBIS$Q34, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q35 <- as.factor(sgbar.likert_ROBIS$Q35)
sgbar.likert_ROBIS$Q35 <- revalue(sgbar.likert_ROBIS$Q35, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q41 <- as.factor(sgbar.likert_ROBIS$Q41)
sgbar.likert_ROBIS$Q41 <- revalue(sgbar.likert_ROBIS$Q41, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q42 <- as.factor(sgbar.likert_ROBIS$Q42)
sgbar.likert_ROBIS$Q42 <- revalue(sgbar.likert_ROBIS$Q42, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q43 <- as.factor(sgbar.likert_ROBIS$Q43)
sgbar.likert_ROBIS$Q43 <- revalue(sgbar.likert_ROBIS$Q43, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))
```

```
sgbar.likert_ROBIS$Q44 <- as.factor(sgbar.likert_ROBIS$Q44)
sgbar.likert_ROBIS$Q44 <- revalue(sgbar.likert_ROBIS$Q44, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q45 <- as.factor(sgbar.likert_ROBIS$Q45)
sgbar.likert_ROBIS$Q45 <- revalue(sgbar.likert_ROBIS$Q45, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))

sgbar.likert_ROBIS$Q46 <- as.factor(sgbar.likert_ROBIS$Q46)
sgbar.likert_ROBIS$Q46 <- revalue(sgbar.likert_ROBIS$Q46, c("0"="Low", "1"="Probably
low", "2"="Probably high", "3"="High"))


######## plotting Likerts scales ----------------
##### likert scale AMSTAR

plot(likert(sgbar.likert[1:11],
            grouping=sgbar.likert$RoB_ROBIS),
            main="",
            ylab="AMSTAR items")

##### likert scale ROBIS

plot(likert(sgbar.likert_ROBIS[1:21],
            grouping=sgbar.likert_ROBIS$AMSTAR_levels_2),
            main="",
            ylab="ROBIS items")

##### ONLY high quality subgroup Likert scales

DB2_high<-subset(DB2, AMSTAR_levels_2=="high_quality")
DB2_high_robis_H<-subset(DB2_high, RoB_ROBIS=="HIGH")
DB2_high_robis_L<-subset(DB2_high, RoB_ROBIS=="LOW")

### ROBIS HIGH

sgbar.likert_ROBIS_high_H <- DB2_high_robis_H[,c(4:24)]
desired.order <- c("Low", "Probably low","Probably high", "High")
sgbar.likert_ROBIS_high_H$Q11 <- as.factor(sgbar.likert_ROBIS_high_H$Q11)
sgbar.likert_ROBIS_high_H$Q11 <- revalue(sgbar.likert_ROBIS_high_H$Q11, c("0"="Low",
"1"="Probably low","2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q11 <- factor(sgbar.likert_ROBIS_high_H$Q11, levels=desired.order,
ordered=TRUE)
```

```
sgbar.likert_ROBIS_high_H$Q12 <- as.factor(sgbar.likert_ROBIS_high_H$Q12)
sgbar.likert_ROBIS_high_H$Q12 <- revalue(sgbar.likert_ROBIS_high_H$Q12, c("0"="Low","1"="Probably
low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q12 <- factor(sgbar.likert_ROBIS_high_H$Q12, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q13 <- as.factor(sgbar.likert_ROBIS_high_H$Q13)
sgbar.likert_ROBIS_high_H$Q13 <- revalue(sgbar.likert_ROBIS_high_H$Q13, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q13 <- factor(sgbar.likert_ROBIS_high_H$Q13, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q14 <- as.factor(sgbar.likert_ROBIS_high_H$Q14)
sgbar.likert_ROBIS_high_H$Q14 <- revalue(sgbar.likert_ROBIS_high_H$Q14, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q14 <- factor(sgbar.likert_ROBIS_high_H$Q14, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q15 <- as.factor(sgbar.likert_ROBIS_high_H$Q15)
sgbar.likert_ROBIS_high_H$Q15 <- revalue(sgbar.likert_ROBIS_high_H$Q15, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q15 <- factor(sgbar.likert_ROBIS_high_H$Q15, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q21 <- as.factor(sgbar.likert_ROBIS_high_H$Q21)
sgbar.likert_ROBIS_high_H$Q21 <- revalue(sgbar.likert_ROBIS_high_H$Q21, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q21 <- factor(sgbar.likert_ROBIS_high_H$Q21, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q22 <- as.factor(sgbar.likert_ROBIS_high_H$Q22)
sgbar.likert_ROBIS_high_H$Q22 <- revalue(sgbar.likert_ROBIS_high_H$Q22, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q22 <- factor(sgbar.likert_ROBIS_high_H$Q22, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q23 <- as.factor(sgbar.likert_ROBIS_high_H$Q23)
sgbar.likert_ROBIS_high_H$Q23 <- revalue(sgbar.likert_ROBIS_high_H$Q23, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q23 <- factor(sgbar.likert_ROBIS_high_H$Q23, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q24 <- as.factor(sgbar.likert_ROBIS_high_H$Q24)
```

```
sgbar.likert_ROBIS_high_H$Q24 <- revalue(sgbar.likert_ROBIS_high_H$Q24, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q24 <- factor(sgbar.likert_ROBIS_high_H$Q24, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q25 <- as.factor(sgbar.likert_ROBIS_high_H$Q25)
sgbar.likert_ROBIS_high_H$Q25 <- revalue(sgbar.likert_ROBIS_high_H$Q25, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q25 <- factor(sgbar.likert_ROBIS_high_H$Q25, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q31 <- as.factor(sgbar.likert_ROBIS_high_H$Q31)
sgbar.likert_ROBIS_high_H$Q31 <- revalue(sgbar.likert_ROBIS_high_H$Q31, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q31 <- factor(sgbar.likert_ROBIS_high_H$Q31, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q32 <- as.factor(sgbar.likert_ROBIS_high_H$Q32)
sgbar.likert_ROBIS_high_H$Q32 <- revalue(sgbar.likert_ROBIS_high_H$Q32, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q32 <- factor(sgbar.likert_ROBIS_high_H$Q32, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q33 <- as.factor(sgbar.likert_ROBIS_high_H$Q33)
sgbar.likert_ROBIS_high_H$Q33 <- revalue(sgbar.likert_ROBIS_high_H$Q33, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q33 <- factor(sgbar.likert_ROBIS_high_H$Q33, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q34 <- as.factor(sgbar.likert_ROBIS_high_H$Q34)
sgbar.likert_ROBIS_high_H$Q34 <- revalue(sgbar.likert_ROBIS_high_H$Q34, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q34 <- factor(sgbar.likert_ROBIS_high_H$Q34, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q35 <- as.factor(sgbar.likert_ROBIS_high_H$Q35)
sgbar.likert_ROBIS_high_H$Q35 <- revalue(sgbar.likert_ROBIS_high_H$Q35, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q35 <- factor(sgbar.likert_ROBIS_high_H$Q35, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q41 <- as.factor(sgbar.likert_ROBIS_high_H$Q41)
sgbar.likert_ROBIS_high_H$Q41 <- revalue(sgbar.likert_ROBIS_high_H$Q41, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
```

```
sgbar.likert_ROBIS_high_H$Q41 <- factor(sgbar.likert_ROBIS_high_H$Q41, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q42 <- as.factor(sgbar.likert_ROBIS_high_H$Q42)
sgbar.likert_ROBIS_high_H$Q42 <- revalue(sgbar.likert_ROBIS_high_H$Q42, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q42 <- factor(sgbar.likert_ROBIS_high_H$Q42, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q43 <- as.factor(sgbar.likert_ROBIS_high_H$Q43)
sgbar.likert_ROBIS_high_H$Q43 <- revalue(sgbar.likert_ROBIS_high_H$Q43, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q43 <- factor(sgbar.likert_ROBIS_high_H$Q43, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q44 <- as.factor(sgbar.likert_ROBIS_high_H$Q44)
sgbar.likert_ROBIS_high_H$Q44 <- revalue(sgbar.likert_ROBIS_high_H$Q44, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q44 <- factor(sgbar.likert_ROBIS_high_H$Q44, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q45 <- as.factor(sgbar.likert_ROBIS_high_H$Q45)
sgbar.likert_ROBIS_high_H$Q45 <- revalue(sgbar.likert_ROBIS_high_H$Q45, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q45 <- factor(sgbar.likert_ROBIS_high_H$Q45, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_H$Q46 <- as.factor(sgbar.likert_ROBIS_high_H$Q46)
sgbar.likert_ROBIS_high_H$Q46 <- revalue(sgbar.likert_ROBIS_high_H$Q46, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_H$Q46 <- factor(sgbar.likert_ROBIS_high_H$Q46, levels=desired.order,
ordered=TRUE)

attach(sgbar.likert_ROBIS_high_H)

######## likert scale ROBIS subgrupo high ----------------

plot(likert(sgbar.likert_ROBIS_high_H[1:21]),
     main="High RoB (n=19)",
     ylab="ROBIS items")


######## ROBIS LOW ----------------
```

```
sgbar.likert_ROBIS_high_L <- DB2_high_robis_L[,c(4:24)]
desired.order <- c("Low", "Probably low","Probably high", "High")
sgbar.likert_ROBIS_high_L$Q11 <- as.factor(sgbar.likert_ROBIS_high_L$Q11)
sgbar.likert_ROBIS_high_L$Q11 <- revalue(sgbar.likert_ROBIS_high_L$Q11, c("0"="Low",
"1"="Probably low","2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q11 <- factor(sgbar.likert_ROBIS_high_L$Q11, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q12 <- as.factor(sgbar.likert_ROBIS_high_L$Q12)
sgbar.likert_ROBIS_high_L$Q12 <- revalue(sgbar.likert_ROBIS_high_L$Q12, c("0"="Low","1"="Probably
low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q12 <- factor(sgbar.likert_ROBIS_high_L$Q12, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q13 <- as.factor(sgbar.likert_ROBIS_high_L$Q13)
sgbar.likert_ROBIS_high_L$Q13 <- revalue(sgbar.likert_ROBIS_high_L$Q13, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q13 <- factor(sgbar.likert_ROBIS_high_L$Q13, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q14 <- as.factor(sgbar.likert_ROBIS_high_L$Q14)
sgbar.likert_ROBIS_high_L$Q14 <- revalue(sgbar.likert_ROBIS_high_L$Q14, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q14 <- factor(sgbar.likert_ROBIS_high_L$Q14, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q15 <- as.factor(sgbar.likert_ROBIS_high_L$Q15)
sgbar.likert_ROBIS_high_L$Q15 <- revalue(sgbar.likert_ROBIS_high_L$Q15, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q15 <- factor(sgbar.likert_ROBIS_high_L$Q15, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q21 <- as.factor(sgbar.likert_ROBIS_high_L$Q21)
sgbar.likert_ROBIS_high_L$Q21 <- revalue(sgbar.likert_ROBIS_high_L$Q21, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q21 <- factor(sgbar.likert_ROBIS_high_L$Q21, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q22 <- as.factor(sgbar.likert_ROBIS_high_L$Q22)
sgbar.likert_ROBIS_high_L$Q22 <- revalue(sgbar.likert_ROBIS_high_L$Q22, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q22 <- factor(sgbar.likert_ROBIS_high_L$Q22, levels=desired.order,
ordered=TRUE)
```

```
sgbar.likert_ROBIS_high_L$Q23 <- as.factor(sgbar.likert_ROBIS_high_L$Q23)
sgbar.likert_ROBIS_high_L$Q23 <- revalue(sgbar.likert_ROBIS_high_L$Q23, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q23 <- factor(sgbar.likert_ROBIS_high_L$Q23, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q24 <- as.factor(sgbar.likert_ROBIS_high_L$Q24)
sgbar.likert_ROBIS_high_L$Q24 <- revalue(sgbar.likert_ROBIS_high_L$Q24, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q24 <- factor(sgbar.likert_ROBIS_high_L$Q24, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q25 <- as.factor(sgbar.likert_ROBIS_high_L$Q25)
sgbar.likert_ROBIS_high_L$Q25 <- revalue(sgbar.likert_ROBIS_high_L$Q25, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q25 <- factor(sgbar.likert_ROBIS_high_L$Q25, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q31 <- as.factor(sgbar.likert_ROBIS_high_L$Q31)
sgbar.likert_ROBIS_high_L$Q31 <- revalue(sgbar.likert_ROBIS_high_L$Q31, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q31 <- factor(sgbar.likert_ROBIS_high_L$Q31, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q32 <- as.factor(sgbar.likert_ROBIS_high_L$Q32)
sgbar.likert_ROBIS_high_L$Q32 <- revalue(sgbar.likert_ROBIS_high_L$Q32, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q32 <- factor(sgbar.likert_ROBIS_high_L$Q32, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q33 <- as.factor(sgbar.likert_ROBIS_high_L$Q33)
sgbar.likert_ROBIS_high_L$Q33 <- revalue(sgbar.likert_ROBIS_high_L$Q33, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q33 <- factor(sgbar.likert_ROBIS_high_L$Q33, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q34 <- as.factor(sgbar.likert_ROBIS_high_L$Q34)
sgbar.likert_ROBIS_high_L$Q34 <- revalue(sgbar.likert_ROBIS_high_L$Q34, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q34 <- factor(sgbar.likert_ROBIS_high_L$Q34, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q35 <- as.factor(sgbar.likert_ROBIS_high_L$Q35)
```

```
sgbar.likert_ROBIS_high_L$Q35 <- revalue(sgbar.likert_ROBIS_high_L$Q35, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q35 <- factor(sgbar.likert_ROBIS_high_L$Q35, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q41 <- as.factor(sgbar.likert_ROBIS_high_L$Q41)
sgbar.likert_ROBIS_high_L$Q41 <- revalue(sgbar.likert_ROBIS_high_L$Q41, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q41 <- factor(sgbar.likert_ROBIS_high_L$Q41, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q42 <- as.factor(sgbar.likert_ROBIS_high_L$Q42)
sgbar.likert_ROBIS_high_L$Q42 <- revalue(sgbar.likert_ROBIS_high_L$Q42, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q42 <- factor(sgbar.likert_ROBIS_high_L$Q42, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q43 <- as.factor(sgbar.likert_ROBIS_high_L$Q43)
sgbar.likert_ROBIS_high_L$Q43 <- revalue(sgbar.likert_ROBIS_high_L$Q43, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q43 <- factor(sgbar.likert_ROBIS_high_L$Q43, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q44 <- as.factor(sgbar.likert_ROBIS_high_L$Q44)
sgbar.likert_ROBIS_high_L$Q44 <- revalue(sgbar.likert_ROBIS_high_L$Q44, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q44 <- factor(sgbar.likert_ROBIS_high_L$Q44, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q45 <- as.factor(sgbar.likert_ROBIS_high_L$Q45)
sgbar.likert_ROBIS_high_L$Q45 <- revalue(sgbar.likert_ROBIS_high_L$Q45, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q45 <- factor(sgbar.likert_ROBIS_high_L$Q45, levels=desired.order,
ordered=TRUE)

sgbar.likert_ROBIS_high_L$Q46 <- as.factor(sgbar.likert_ROBIS_high_L$Q46)
sgbar.likert_ROBIS_high_L$Q46 <- revalue(sgbar.likert_ROBIS_high_L$Q46, c("0"="Low",
"1"="Probably low", "2"="Probably high", "3"="High"))
sgbar.likert_ROBIS_high_L$Q46 <- factor(sgbar.likert_ROBIS_high_L$Q46, levels=desired.order,
ordered=TRUE)

attach(sgbar.likert_ROBIS_high_L)

########  Likert scale ROBIS high soubgroup --------------------
```

```
plot(likert(sgbar.likert_ROBIS_high_L[1:21]), main="Low RoB (n=12)", ylab="ROBIS items")
b<-likert(sgbar.likert_ROBIS_high[1:21])


########  ROBIS ALL --------------------

DB4_all<- DB4[,c(12:32)]
desired.order <- c("Low", "Probably low","Probably high", "High")
DB4_all$Q11 <- as.factor(DB4_all$Q11)
DB4_all$Q11 <- revalue(DB4_all$Q11, c("0"="Low", "1"="Probably low","2"="Probably high",
"3"="High"))
DB4_all$Q11 <- factor(DB4_all$Q11, levels=desired.order, ordered=TRUE)

DB4_all$Q12 <- as.factor(DB4_all$Q12)
DB4_all$Q12 <- revalue(DB4_all$Q12, c("0"="Low","1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q12 <- factor(DB4_all$Q12, levels=desired.order, ordered=TRUE)

DB4_all$Q13 <- as.factor(DB4_all$Q13)
DB4_all$Q13 <- revalue(DB4_all$Q13, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q13 <- factor(DB4_all$Q13, levels=desired.order, ordered=TRUE)

DB4_all$Q14 <- as.factor(DB4_all$Q14)
DB4_all$Q14 <- revalue(DB4_all$Q14, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q14 <- factor(DB4_all$Q14, levels=desired.order, ordered=TRUE)

DB4_all$Q15 <- as.factor(DB4_all$Q15)
DB4_all$Q15 <- revalue(DB4_all$Q15, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q15 <- factor(DB4_all$Q15, levels=desired.order, ordered=TRUE)

DB4_all$Q21 <- as.factor(DB4_all$Q21)
DB4_all$Q21 <- revalue(DB4_all$Q21, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q21 <- factor(DB4_all$Q21, levels=desired.order, ordered=TRUE)

DB4_all$Q22 <- as.factor(DB4_all$Q22)
DB4_all$Q22 <- revalue(DB4_all$Q22, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q22 <- factor(DB4_all$Q22, levels=desired.order, ordered=TRUE)
```

```
DB4_all$Q23 <- as.factor(DB4_all$Q23)
DB4_all$Q23 <- revalue(DB4_all$Q23, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q23 <- factor(DB4_all$Q23, levels=desired.order, ordered=TRUE)

DB4_all$Q24 <- as.factor(DB4_all$Q24)
DB4_all$Q24 <- revalue(DB4_all$Q24, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q24 <- factor(DB4_all$Q24, levels=desired.order, ordered=TRUE)

DB4_all$Q25 <- as.factor(DB4_all$Q25)
DB4_all$Q25 <- revalue(DB4_all$Q25, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q25 <- factor(DB4_all$Q25, levels=desired.order, ordered=TRUE)

DB4_all$Q31 <- as.factor(DB4_all$Q31)
DB4_all$Q31 <- revalue(DB4_all$Q31, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q31 <- factor(DB4_all$Q31, levels=desired.order, ordered=TRUE)

DB4_all$Q32 <- as.factor(DB4_all$Q32)
DB4_all$Q32 <- revalue(DB4_all$Q32, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q32 <- factor(DB4_all$Q32, levels=desired.order, ordered=TRUE)

DB4_all$Q33 <- as.factor(DB4_all$Q33)
DB4_all$Q33 <- revalue(DB4_all$Q33, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q33 <- factor(DB4_all$Q33, levels=desired.order, ordered=TRUE)

DB4_all$Q34 <- as.factor(DB4_all$Q34)
DB4_all$Q34 <- revalue(DB4_all$Q34, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q34 <- factor(DB4_all$Q34, levels=desired.order, ordered=TRUE)

DB4_all$Q35 <- as.factor(DB4_all$Q35)
DB4_all$Q35 <- revalue(DB4_all$Q35, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q35 <- factor(DB4_all$Q35, levels=desired.order, ordered=TRUE)

DB4_all$Q41 <- as.factor(DB4_all$Q41)
DB4_all$Q41 <- revalue(DB4_all$Q41, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q41 <- factor(DB4_all$Q41, levels=desired.order, ordered=TRUE)
```

```
DB4_all$Q42 <- as.factor(DB4_all$Q42)
DB4_all$Q42 <- revalue(DB4_all$Q42, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q42 <- factor(DB4_all$Q42, levels=desired.order, ordered=TRUE)

DB4_all$Q43 <- as.factor(DB4_all$Q43)
DB4_all$Q43 <- revalue(DB4_all$Q43, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q43 <- factor(DB4_all$Q43, levels=desired.order, ordered=TRUE)

DB4_all$Q44 <- as.factor(DB4_all$Q44)
DB4_all$Q44 <- revalue(DB4_all$Q44, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q44 <- factor(DB4_all$Q44, levels=desired.order, ordered=TRUE)

DB4_all$Q45 <- as.factor(DB4_all$Q45)
DB4_all$Q45 <- revalue(DB4_all$Q45, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q45 <- factor(DB4_all$Q45, levels=desired.order, ordered=TRUE)

DB4_all$Q46 <- as.factor(DB4_all$Q46)
DB4_all$Q46 <- revalue(DB4_all$Q46, c("0"="Low", "1"="Probably low", "2"="Probably high",
"3"="High"))
DB4_all$Q46 <- factor(DB4_all$Q46, levels=desired.order, ordered=TRUE)

attach(DB4_all)

##### likert scale ROBIS subgrupo high

plot(likert(DB4_all[1:21]),
     main="All SRs (n=139)",
     ylab="ROBIS items")
likert.bar.plot(likert(DB4_all[1:21]),
                centered = FALSE,
                main="All SRs (n=139)",
                ylab="ROBIS items",
                legend="",
                legend.position = "right",
                ordered=FALSE,
                low.color="lightsalmon1",
                high.color="skyblue3",
                neutral.color="seagreen3")
```

```
########### with DIMENSIONS

file4<-read.csv2("plot_paper_ROBIS.csv",
                 sep = ";",
                 dec= ".",
                 stringsAsFactors = TRUE,
                 header = TRUE)
DB4<-as.data.frame(file4)
DB4$D1_sum <- (DB4$Q11+DB4$Q12+DB4$Q13+DB4$Q14+DB4$Q15)
DB4$D2_sum <- (DB4$Q21+DB4$Q22+DB4$Q23+DB4$Q24+DB4$Q25)
DB4$D3_sum <- (DB4$Q31+DB4$Q32+DB4$Q33+DB4$Q34+DB4$Q35)
DB4$D4_sum <- (DB4$Q41+DB4$Q42+DB4$Q43+DB4$Q44+DB4$Q45+DB4$Q46)
DB4$radius<-(D1_sum+D2_sum+D3_sum+D4_sum)
desired.order <- c("High","Low", "Unclear")

DB4$D1 <- as.factor(DB4$D1)
DB4$D1 <- revalue(DB4$D1, c("LOW"="Low", "HIGH"="High", "UNCLEAR"="Unclear"))
DB4$D1 <- factor(DB4$D1, levels=desired.order, ordered=TRUE)

DB4$D2 <- as.factor(DB4$D2)
DB4$D2 <- revalue(DB4$D2, c("LOW"="Low", "HIGH"="High", "UNCLEAR"="Unclear"))
DB4$D2 <- factor(DB4$D2, levels=desired.order, ordered=TRUE)

DB4$D3 <- as.factor(DB4$D3)
DB4$D3 <- revalue(DB4$D3, c("LOW"="Low", "HIGH"="High", "UNCLEAR"="Unclear"))
DB4$D3 <- factor(DB4$D3, levels=desired.order, ordered=TRUE)

DB4$D4 <- as.factor(DB4$D4)
DB4$D4 <- revalue(DB4$D4, c("LOW"="Low", "HIGH"="High", "UNCLEAR"="Unclear"))
DB4$D4 <- factor(DB4$D4, levels=desired.order, ordered=TRUE)

DB4$RoB <- as.factor(DB4$RoB)
DB4$RoB <- revalue(DB4$RoB, c("LOW"="Low", "HIGH"="High", "UNCLEAR"="Unclear"))
DB4$RoB <- factor(DB4$RoB, levels=desired.order, ordered=TRUE)

names(DB4)[3] <- "1. Study elegibility criteria"
names(DB4)[4] <- "2. Identification and selection of studies"
names(DB4)[5] <- "3. Data collection and study appraisal"
names(DB4)[6] <- "4. Synthesis and findings"
names(DB4)[7] <- "Risk of Bias in the review"

likert.bar.plot(likert(DB4[3:7]),
                centered = FALSE,
                legend="",
```

```
                legend.position = "right",
                ordered=FALSE,
                low.color="lightsalmon1",
                high.color="skyblue3",
                neutral.color="seagreen3")


#############################################################################
######## 3: correlation matrix of ROBIS items ----------------

corr_robis_melt <- melt(cor(DB2[,4:24]))
corr_robis_melt$value <-round(corr_robis_melt$value, 2)
corr_robis_melt<- subset(corr_robis_melt, value < 1)
ggplot(corr_robis_melt, aes(Var1, Var2)) +
  geom_tile(aes(fill = value)) +
  geom_text(aes(label=value), size = 3, fontface = "bold") +
  #scale_fill_gradient(low = "lightgreen", high = "red") +
  scale_fill_gradientn(colours=c("darkgreen","lightgreen","orange", "red"), na.value
= "grey98", limits = c(-0.16, 1), breaks=seq(min(corr_robis_melt$value), max(corr_robis_melt$valu
by=10)) +
  theme_minimal() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.border = element_blank(),
        panel.background = element_blank(),
        axis.title = element_blank(),
        axis.text = element_text(size = 12, face = "bold"))


#############################################################################
######## 4: correlation network of ROBIS items ----------------

#### packages

library(ggraph)
library(igraph)

#### subsetting coef spearman >0.5

corr_robis_melt_0.6 <- subset(corr_robis_melt, value>0.5)

####  network plotting

graph <- graph_from_data_frame(corr_robis_melt_0.6, directed = FALSE)
```

```
p <- ggraph(graph, layout = 'kk') +
    geom_edge_link() +
    geom_node_point(show.legend = TRUE) +
    ggtitle('') +
    geom_node_text(aes(label = name), repel = TRUE) +
    geom_edge_link(aes(edge_alpha = abs(value), edge_width = abs(value), color = value))
+
    guides(edge_alpha = "none", edge_width = "none") +
    scale_edge_colour_gradientn(limits = c(0.55, 0.9), colors = c("orange","firebrick2"))
+
    geom_node_point(color = "darkblue", size = 3)
p + theme_graph()
dodgerblue2


################################################################################
######### 5: Visualizing Multivariate Data with Radviz ---------------------


library ("Radviz")

#########  ROBIS signaling questions --------------------

######## normalizing the data
norm <- apply(DB2[,4:24],2,do.L,fun=function(x) quantile(x,c(0.025,0.975)))

######## defining the anchors
ct.S <- make.S(dimnames(DB2[,4:24])[[2]])

######## computing the similarity matrix (distances between columns and rows)
ct.sim <- cosine(norm)

######## getting the current radviz-independent measure of projection efficiency
in.da(ct.S,ct.sim)

######## getting the current radviz-dependent measure of projection efficiency
rv.da(ct.S,ct.sim)

######## The radviz-independent score should be maximal when the dimensional anchor positions
are optimal.
######## The radviz-dependent score should be minimal when the dimensional anchor positions
are optimal.
######## Optimization procedure:
```

```
optim.ct <- do.optim(ct.S,ct.sim,iter=100,n=1000)
ct.S <- make.S(tail(optim.ct$best,1)[[1]])

######## Getting final projections -----------------
######## The do.radviz function will then use the normalized values and the Springs
######## to project each thing in a 2D space:
ct.rv <- do.radviz(norm,ct.S)

######## Visualizing the results  ----------------
########  There is a S3 plot function defined for radviz; using the default will give
the following result

#### AMSTAR
sub.rv_low_quality <- subset(ct.rv,DB2$AMSTAR_levels_2=="low_quality")
sub.rv_moderate_quality <- subset(ct.rv,DB2$AMSTAR_levels_2=="moderate_quality")
sub.rv_high_quality <- subset(ct.rv,DB2$AMSTAR_levels_2=="high_quality")
plot(ct.rv,point.shape=19, point.color=c("coral", "seagreen3", "dodgerblue2")[DB2$AMSTAR_levels_2]
contour(sub.rv_low_quality, add=T, contour.color = "dodgerblue2")
contour(sub.rv_moderate_quality, add=T, contour.color = "seagreen3")
contour(sub.rv_high_quality, add=T, contour.color = "coral")

#### ROBIS
sub.rv_low <- subset(ct.rv,DB2$RoB_ROBIS=="LOW")
sub.rv_high <- subset(ct.rv,DB2$RoB_ROBIS=="HIGH")
plot(ct.rv,point.shape=19, point.color=c("coral", "turquoise3")[DB2$RoB_ROBIS])
contour(sub.rv_low, add=T, contour.color = "turquoise3")
contour(sub.rv_high, add=T, contour.color = "coral")

####  ROBIS domains
####  normalizing the data
DB2$D1<-DB2[,4]+DB2[,5]+DB2[,6]+DB2[,7]+DB2[,8]
DB2$D2<-DB2[,9]+DB2[,10]+DB2[,11]+DB2[,12]+DB2[,13]
DB2$D3<-DB2[,14]+DB2[,15]+DB2[,16]+DB2[,17]+DB2[,18]
DB2$D4<-DB2[,19]+DB2[,20]+DB2[,21]+DB2[,22]+DB2[,23]+DB2[,24]
attach(DB2)
norm <- apply(DB2[,43:46],2,do.L,fun=function(x) quantile(x,c(0.025,0.975)))

ct.S <- make.S(dimnames(DB2[,43:46])[[2]])
ct.sim <- cosine(norm)
in.da(ct.S,ct.sim)
rv.da(ct.S,ct.sim)
optim.ct <- do.optim(ct.S,ct.sim,iter=100,n=1000)
ct.S <- make.S(tail(optim.ct$best,1)[[1]])
ct.rv <- do.radviz(norm,ct.S)
```

```
# Visualizing the results:
## ROBIS
sub.rv_low <- subset(ct.rv,DB2$RoB_ROBIS=="LOW")
sub.rv_high <- subset(ct.rv,DB2$RoB_ROBIS=="HIGH")
plot(ct.rv,point.shape=19, point.color=c("coral", "turquoise3")[DB2$RoB_ROBIS])
contour(sub.rv_low, add=T, contour.color = "turquoise3")
contour(sub.rv_high, add=T, contour.color = "coral")

## AMSTAR
sub.rv_low_quality <- subset(ct.rv,DB2$AMSTAR_levels_2=="low_quality")
sub.rv_moderate_quality <- subset(ct.rv,DB2$AMSTAR_levels_2=="moderate_quality")
sub.rv_high_quality <- subset(ct.rv,DB2$AMSTAR_levels_2=="high_quality")
plot(ct.rv,point.shape=19, point.color=c("coral", "seagreen3", "dodgerblue2")[DB2$AMSTAR_levels_2])
contour(sub.rv_low_quality, add=T, contour.color = "dodgerblue2")
contour(sub.rv_moderate_quality, add=T, contour.color = "seagreen3")
contour(sub.rv_high_quality, add=T, contour.color = "coral")
```

Apéndice C: Guía para el uso de ROBIS

ROBIS evalúa de forma estructurada la pertinencia de la pregunta de investigación y el riesgo de sesgo. Requiere del conocimiento de conceptos metodológicos y específicos del tema de revisión. Se recomienda que sea realizado por dos revisores independientes o, al menos, uno supervisado. Puede aplicarse a revisiones sietmáticas sobre distintos temas; en este sentido, todas las preguntas de señalización deben ser relevantes. La herramienta implica hacer juicios y fomenta la transparencia al tener que aportar la información que los apoya, la señalización de preguntas y la justificación de los juicios de preocupación general.

En función del objetivo pueden realizarse dos tipos de evaluaciones:

- Si el objetivo es proporcionar una evaluación del riesgo de sesgo en la revisión sistemática con las razones que lo justifican, se requerirá una evaluación completa en la que todas las preguntas de señalización de todos los dominios se evalúan.

- Si el objetivo es identificar si la revisión sistemática está en alto riesgo de sesgo global o si hay preocupaciones con dominios particulares, entonces los evaluadores pueden optar por detenerla una vez que se ha identificado una alta preocupación de riesgo de sesgo.

Consideraciones específicas del uso de ROBIS

La herramienta se completa en tres fases. Las preguntas de señalización ayudan a evaluar las cuestiones específicas sobre posibles sesgos y juzgar el riesgo de sesgo general.

	Phase 2				Phase 3
	1. Study eligibility criteria	**2. Identification and selection of studies**	**3. Data collection and study appraisal**	**4. Synthesis and findings**	**Risk of bias in the review**
Signaling questions	1.1 Did the review adhere to predefined objectives and eligibility criteria?	2.1 Did the search include an appropriate range of databases/ electronic sources for published and unpublished reports?	3.1. Were efforts made to minimize error in data collection?	4.1. Did the synthesis include all studies that it should?	A. Did the interpretation of findings address all of the concerns identified in domains 1 to 4?
	1.2 Were the eligibility criteria appropriate for the review question?	2.2 Were methods additional to database searching used to identify relevant reports?	3.2. Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	4.2. Were all predefined analyses reported or departures explained?	B. Was the relevance of identified studies to the review's research question appropriately considered?
	1.3 Were eligibility criteria unambiguous?	2.3 Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	3.3. Were all relevant study results collected for use in the synthesis?	4.3. Was the synthesis appropriate given the nature and similarity in the research questions, study designs, and outcomes across included studies?	C. Did the reviewers avoid emphasizing results on the basis of their statistical significance?
	1.4 Were all restrictions in eligibility criteria based on study characteristics appropriate?	2.4 Were restrictions based on date, publication format, or language appropriate?	3.4. Was risk of bias (or methodologic quality) formally assessed using appropriate criteria?	4.4. Was between-study variation minimal or addressed in the synthesis?	
	1.5 Were any restrictions in eligibility criteria based on sources of information appropriate?	2.5 Were efforts made to minimize error in selection of studies?	3.5. Were efforts made to minimize error in risk of bias assessment?	4.5. Were the findings robust, for example, as demonstrated through funnel plot or sensitivity analyses? 4.6. Were biases in primary studies minimal or addressed in the synthesis?	
Judgment	Concerns regarding specification of study eligibility criteria	Concerns regarding methods used to identify and/or select studies	Concerns regarding methods used to collect data and appraise studies	Concerns regarding the synthesis	Risk of bias in the review

Fig. 4 Tabla con las preguntas de señalización de los diferentes dominios incluidos en la fase II de ROBIS.

- **Fase I: Evaluación de la pertinencia de la pregunta de investigación (optativa).**
  - En esta fase se informa la pregunta que se está tratando de responder y los evaluadores completan la estrategia definida para las categorías de revisiones sistemáticas. Posteriormente se comparan la pregunta objetivo y la pregunta de la revisión sistemática.
    - Si existe coincidencia puede clasificarse como : Si.
    - Si una o más de las categorías no coinciden, debe ser clasificado como "no".
    - Si hay una coincidencia parcial entre categorías, debería ser calificado como parcial.
- **Fase II: Identificación de incertidumbres sobre sesgos en el proceso de revisión.**
  - Tiene como objetivo identificar las áreas de preocupación sobre el potencial riesgo de sesgo de la revisión sistemática que será juzgado en la evaluación final (fase 3). La evaluación se estructura en dominios que evalúa los procesos clave de la revisión sistemática y deben ser considerados secuencialmente. Dichos dominios son:

* **Dominio 1: Criterios de elegibilidad del estudio.** Evalúa si los criterios de elegibilidad se establecieron a priori de forma clara y apropiada. Por lo general, sólo es posible hacerlo si han sido predefinidos y se dispone de un protocolo o documento de registro previo al desarrollo de la RS. Cuando no se dispone, los evaluadores deben basar su juicio en el informe de los resultados de la revisión sistemática.Una información más detallada puede encontrarse en la tabla 8.
* **Dominio 2: Identificación y selección de estudios.** Este dominio tiene dos objetivos:
  - Evaluar si los estudios primarios que cumplen los criterios de inclusión fueron incluidos en la revisión sistemática. La selección imparcial de los estudios ayuda a garantizar que los que son relevantes se incluyen en la revisión sistemática. Es necesario tener conocimientos metodológicos relacionados con la búsqueda para poder evaluar la sensibilidad de la estrategia además de experiencia en los contenidos relacionados con el tema investigado.
  - El proceso de selección de estudios para su inclusión en la revisión sistemática. Esto implica revisar títulos y resúmenes y evaluar estudios de texto completo para su inclusión. Debe ser llevado a cabo de forma independiente por al menos dos revisores o por uno y revisado por un segundo investigador.Una información más detallada puede encontrarse en la tabla 9.
* **Dominio 3: Extracción de datos y evaluación del riesgo de sesgo de los estudios primarios.** La extracción de los datos incluye su planificación *a priori* y un formulario estructurado que haya sido pilotado. Deben ser recogidos todos los datos que contribuyan a la síntesis e interpretación de los resultados, incluyéndose los datos numéricos y estadísticos y las características generales del estudio primario. Si los datos no están disponibles debe contactarse con los autores. Los errores podrían surgir al transcribirse datos, al no recopilar toda la información relevante o por la subjetividad al recogerlos. La extracción de datos duplicados o extracción única de datos con un control riguroso por un segundo investigador, es esencial para evitar errores aleatorios y posibles sesgos. La validez de los estudios incluidos debe evaluarse utilizando criterios apropiados para su diseño [8]. Deberían realizarla dos revisores, idealmente trabajando independiente-

mente o el segundo verificando las decisiones del primero.Una información más detallada puede encontrarse en la tabla 10.

  - **Dominio 4: Síntesis y hallazgos.** Este dominio evalúa si los revisores han utilizado los métodos apropiados en la síntesis de los datos.Una información más detallada puede encontrarse en la tabla 11. Para algunas RS puede no ser apropiada una síntesis estadística y en su lugar debe realizarse cuantitativa o narrativa. Algunos de los aspectos más importantes a considerar en cualquier síntesis son:
    - Un enfoque analítico apropiado para la pregunta de investigación planteada.
    - El análisis de la heterogeneidad de los datos.
    - La evaluación de los sesgos de los estudios primarios.
    - El análisis del sesgo de publicación.
    - La consideración de sesgos en la notificación de los resultados.

- Cada dominio comprende tres apartados:
  - Información utilizada para apoyar el juicio. Mejora la transparencia y facilita el contraste entre los revisores que completen las evaluaciones de forma independiente.
  - Preguntas de señalización. Referencian una:
    - baja preocupación sobre el riesgo de sesgo si se responden como "sí" o "probablemente sí";
    - una alta preocupación si se responden como "probablemente no" o "no";
    - ausencia de criterio si se responde "ninguna información".
  - Juicio de la preocupación del riesgo de sesgo para cada dominio. Se juzga como "bajo", "alto" o "poco claro". Si las respuestas a todas las preguntas de señalización para un dominio son "sí" o "probablemente sí", el nivel de preocupación puede ser juzgado como bajo. Si se responde a alguna pregunta de señalización "no" o "probablemente no", existe el potencial de preocupación por el sesgo. La categoría de "no información" debe utilizarse sólo cuando se informen datos insuficientes para permitir una sentencia.

- **Fase III: Juicio del riesgo de sesgo en la revisión sistemática.**
  - Esta fase evalúa el riesgo de sesgo de la revisión sistemática en su conjunto. Utiliza la misma estructura que los dominios de la fase 2, incluyendo las pre-

guntas de señalización y la información usada para apoyar el juicio, pero el juicio sobre las preocupaciones sobre sesgo se reemplaza con un juicio general de riesgo de sesgo de la revisión sistemática. Una información más detallada puede encontrarse en la tabla 12.

Presentación de las evaluaciones ROBIS.
Deben resumirse los resultados de la evaluación para todas las RS incluyendo:

- El número de revisiones sistemáticas con preocupación baja, alta o poco clara para cada dominio de la fase 2.
- El número de revisiones sistemáticas con alto o bajo riesgo de sesgo.
- Cuando se utilice, deberá proporcionarse un resumen de la evaluación de la pertinencia.
- Los revisores pueden optar por resaltar las preguntas de señalización en particular, en las que las revisiones sistemáticas siempre se clasifican mal o bien.

Las tablas y las gráficas pueden ser útiles para resumir las evaluaciones de ROBIS en múltiples revisiones sistemáticas. En estos casos, se puede considerar ponderar el número de estudios incluidos o el número total de participantes en cada revisión, en lugar de las revisiones individuales. Alternativamente, los revisores o los desarrolladores de las directrices pueden optar por incluir sólo la revisión sistemática que sea más relevante para su pregunta de destino y con el menor riesgo de sesgo. También se ha sugerido una presentación gráfica para los resultados de una evaluación de ROBIS (cada clasificación de dominio y calificación general) para una sola revisión. ROBIS no debe utilizarse para generar una puntuación total de calidad debido a los problemas asociados con dichas puntuaciones[**?**].

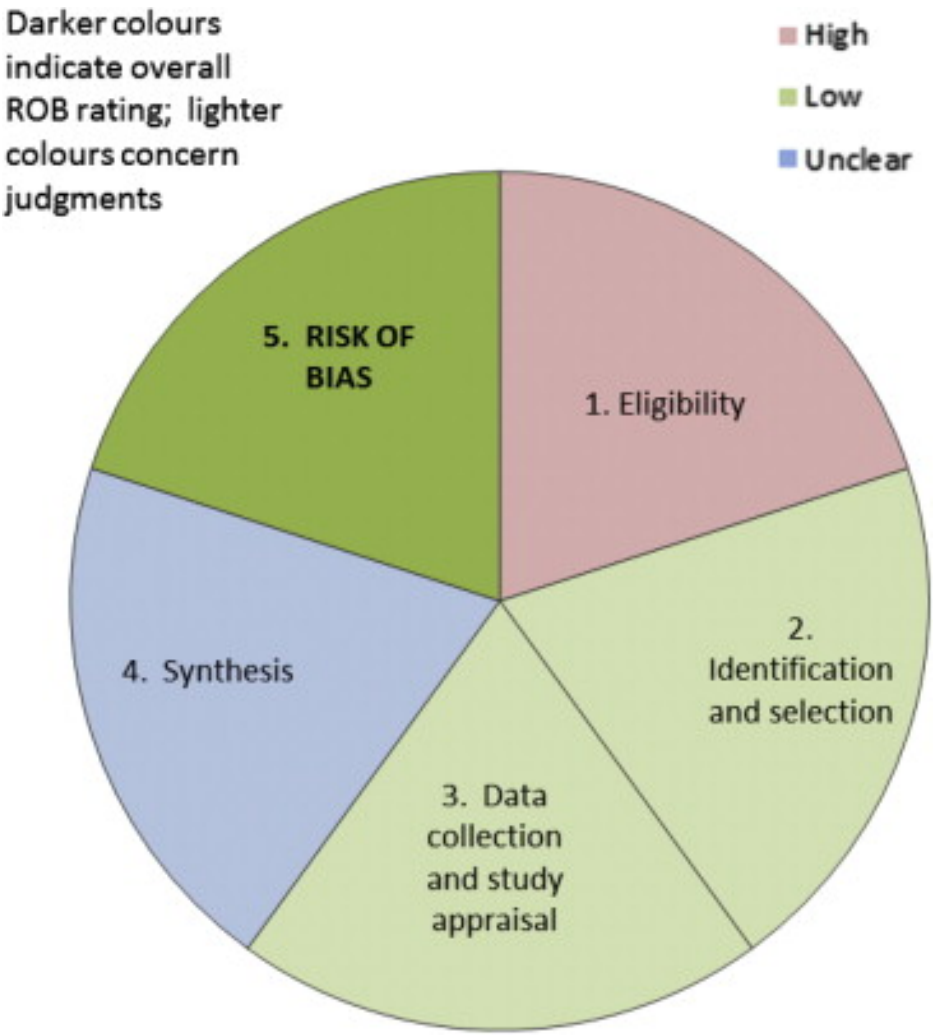


Fig. 5 Representación gráfica del porcentaje de acuerdo con los diferentes dominios de la fase II de ROBIS para una serie de revisiones sistemáticas.

Review	Phase 2				Phase 3
	1. Study eligibility criteria	2. Identification and selection of studies	3. Data collection and study appraisal	4. Synthesis and findings	Risk of bias in the review
1. Perry	☺	☺	☺	?	☺
2. Boehm	☹	☺	☺	☹	☹
3. Mayhew	☺	☹	☹	☺	☺
4. Daya	☹	☹	☹	☹	☹
5. Langhorst	☺	☺	☺	☺	☺
6. Martin-Sanchez	☺	☹	☹	☹	☹
7. Cao	☺	☹	☺	☺	☺
8. Deare	☺	☺	☺	☺	☺
9. Yang	☺	☺	☹	☹	☹
10. Ernst	☹	?	☹	?	?
11. de Sousa Nascimento	☺	☺	☺	☹	☺
12. Holdcroft	☹	☹	☹	?	☹
13. Baronowsky	☹	☹	?	☺	?
14. Terhorst	☺	☹	☺	☹	☹
15. De Silva	☹	☹	☹	?	☺

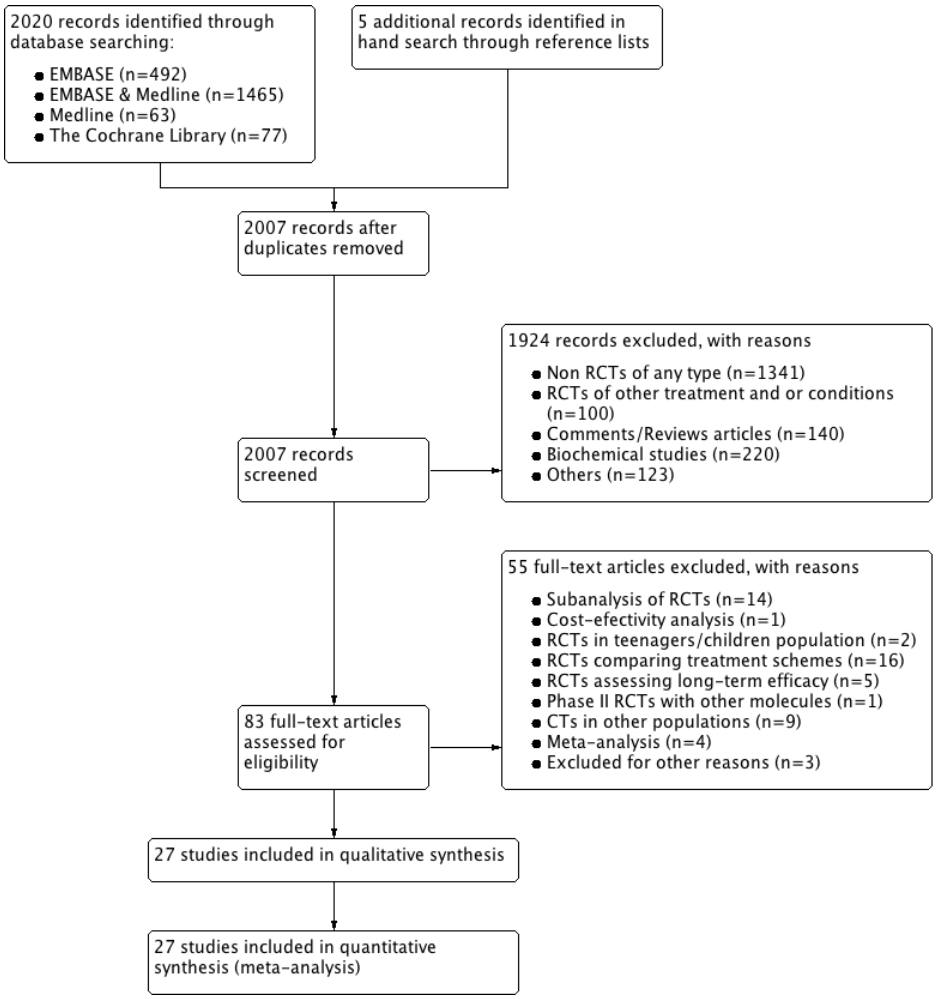
☺ = low risk, ☹ = high risk, ? = unclear risk

Fig. 6 Tabla con el grado de cumplimiento de los dominios de ROBIS de una serie revisiones sistemáticas.

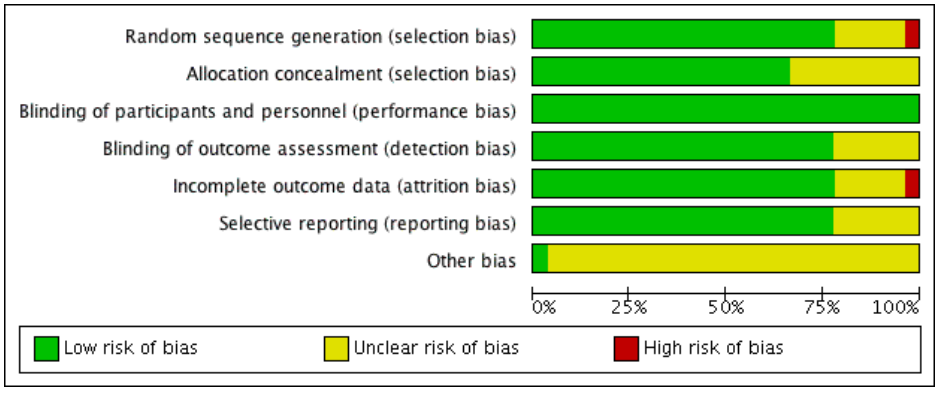
Figuras suplementarias

Artículo 1.

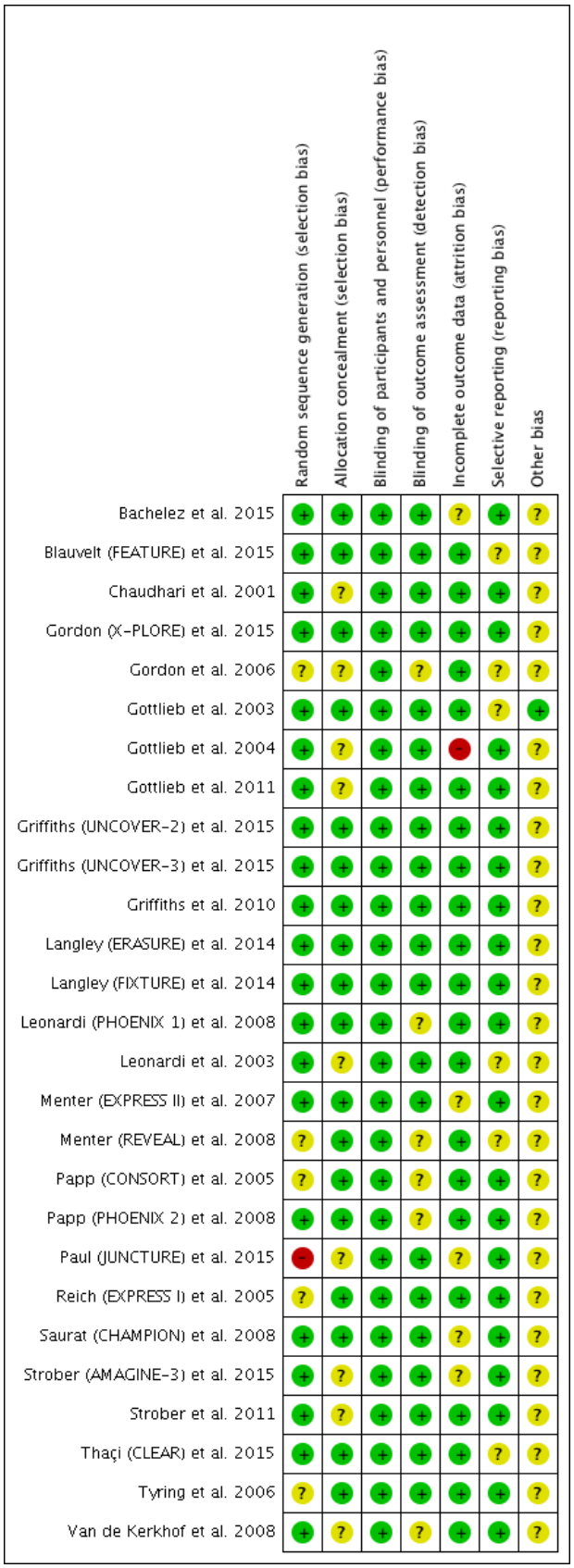
Gómez-García F, Epstein D, Isla-Tejera B, Lorente A, Vélez García-Nieto A, Ruano J. **Short-term efficacy and safety of new biological agents targeting the interleukin-23-T helper 17 pathway for moderate-to-severe plaque psoriasis: a systematic review and network meta-analysis.** *Br J Dermatol* 2017 Mar;176(3):594-603. doi: 10.1111/bjd.14814. Epub 2016 Oct 13.



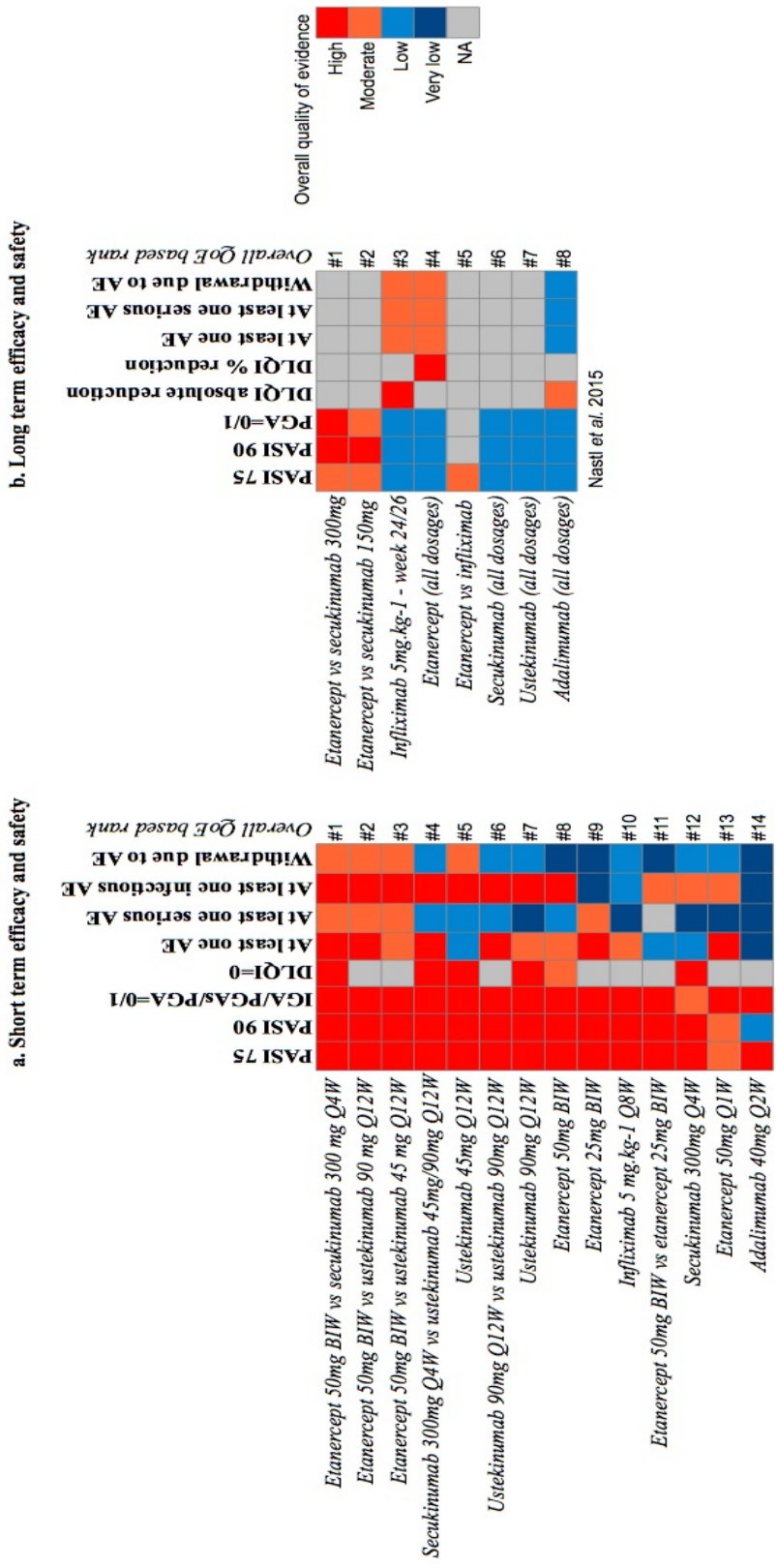
Suppl. Fig. 1: Sample search strategy. This figure shows the process description for selection and elimination of the studies at each stage of the systematic review.



Suppl. Fig. 2: Overall risk of bias.



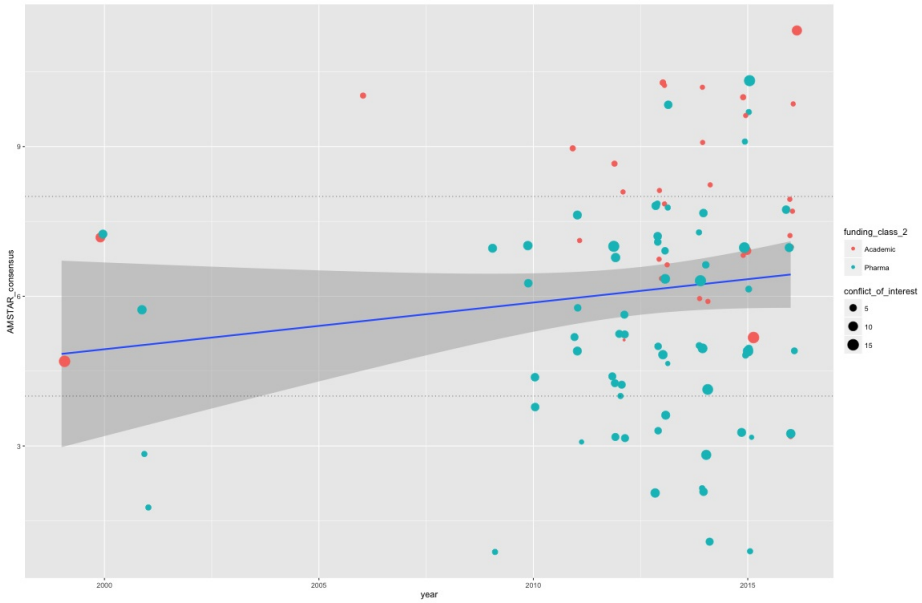
Suppl. Fig. 3: Risk of bias summary.



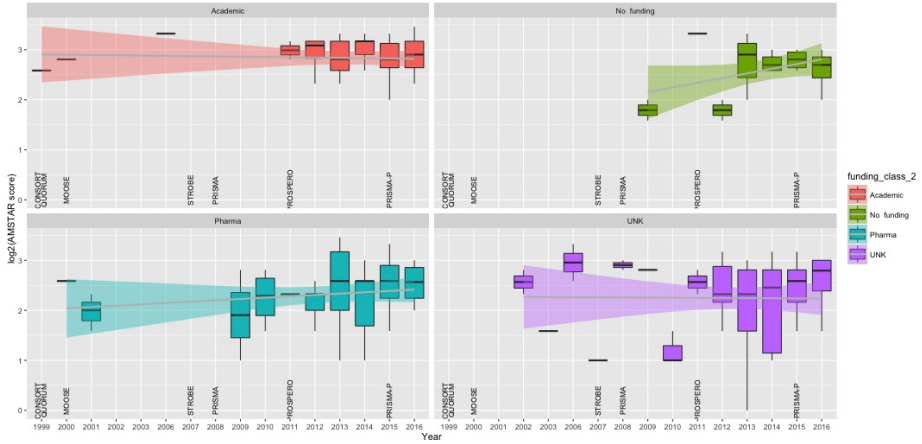
Suppl. Fig. 4: Scoring heatmap of quality of evidence for each outcome across pooled studies.

Artículo 2.

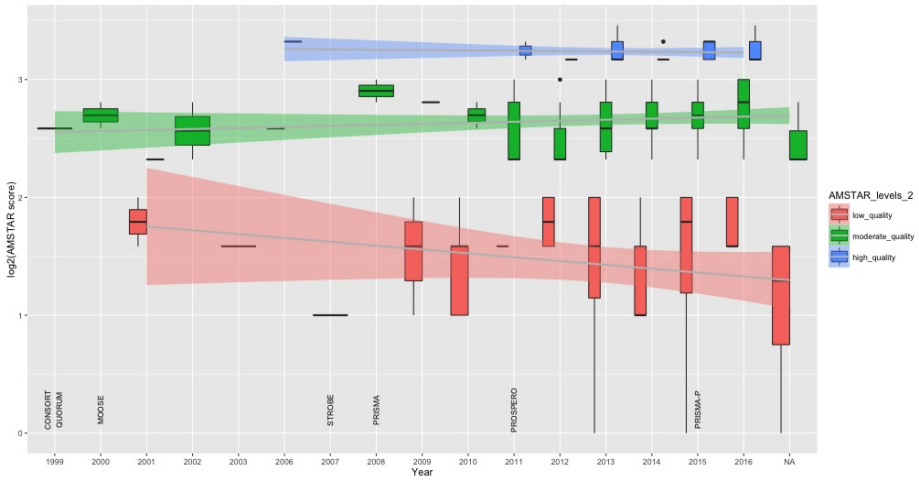
Gómez-García F, Ruano J, Aguilar-Luque M, Gay-Mimbrera J, Maestre-Lopez B, Sanz-Cabanillas JL, Carmona-Fernández PJ, González-Padilla M, Vélez García-Nieto A, Isla-Tejera B. **Systematic reviews and meta-analyses on psoriasis: role of funding sources, conflict of interest and bibliometric indices as predictors of methodological quality.** *Br J Dermatol.* 2017 Jun;176(6):1633-1644. doi: 10.1111/bjd.15380. Epub 2017 May 19.



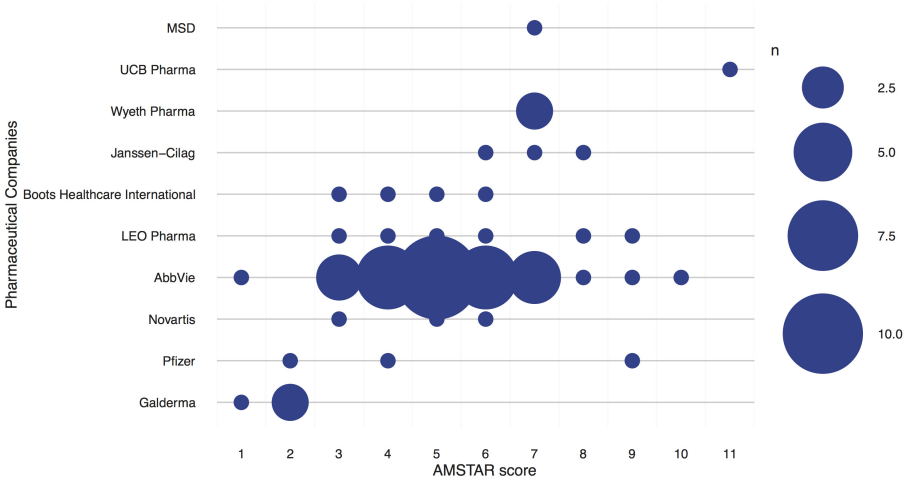
Suppl. Fig. 1: Evolution of AMSTAR scores of SRs and MAs on psoriasis by source of funding and author's conflicts of interest.



Suppl. Fig. 2: Evolution of AMSTAR scores of SRs and MAs on psoriasis by source of funding.



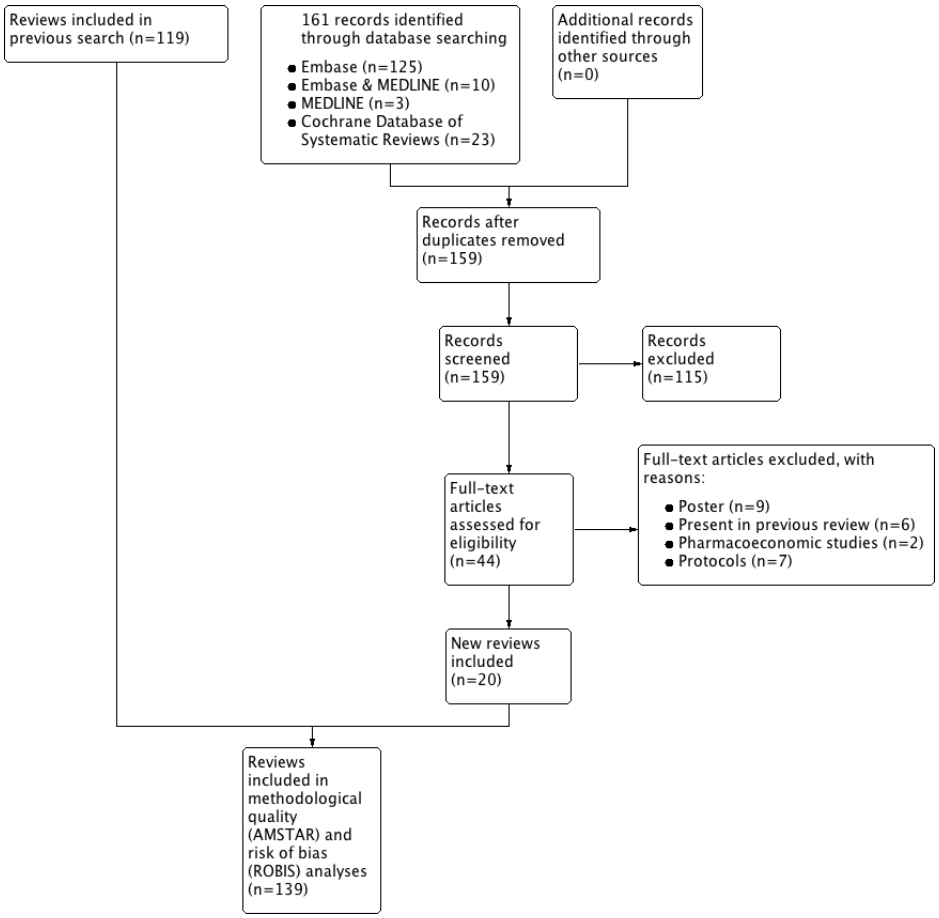
Suppl. Fig. 3: Evolution of AMSTAR scores of SRs and MAs on psoriasis.



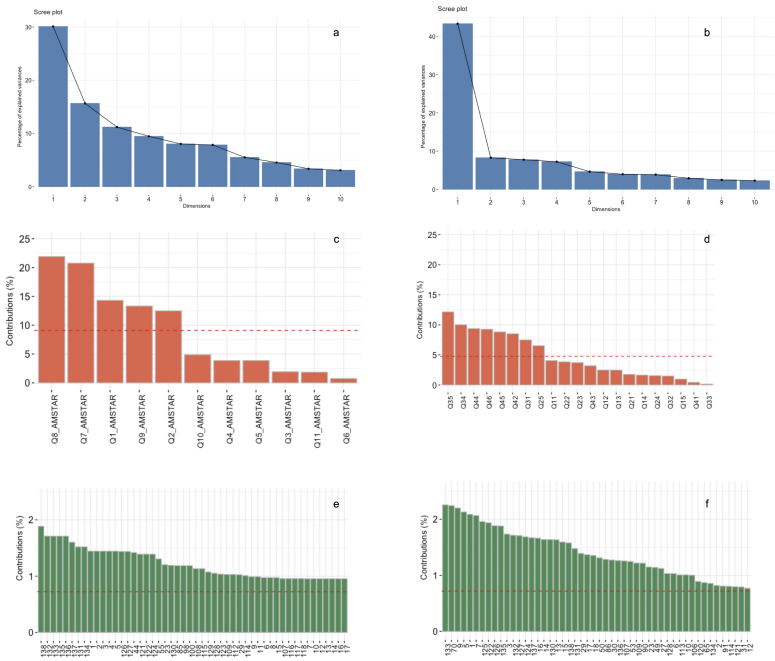
Suppl. Fig. 4: Bubble plot of AMSTAR score distribution of reviews funded by pharmaceutical companies.

Artículo 3.

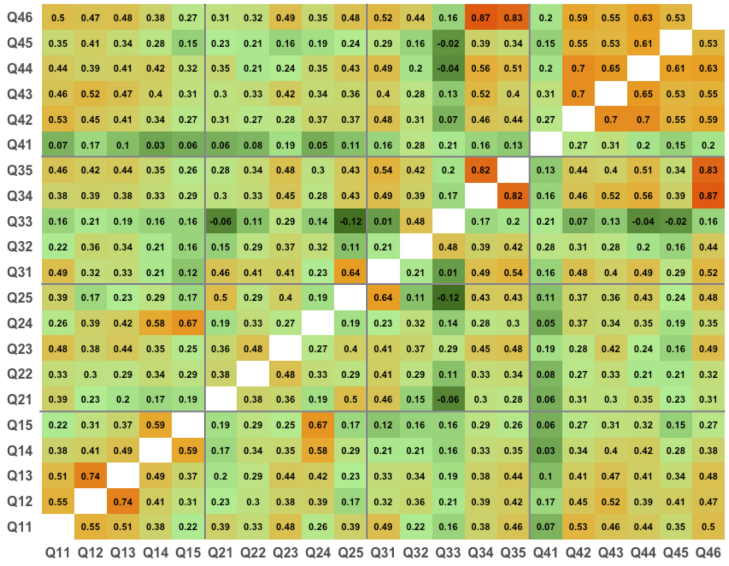
Gómez-García F, Ruano J, Gay-Mimbrera J, Aguilar-Luque M, Sanz-Cabanillas JL, Alcalde-Mellado P, Maestre-López B, Carmona-Fernández PJ, González-Padilla M, García-Nieto AV, Isla-Tejera B. **Most systematic reviews of high methodological quality on psoriasis interventions are classified as high risk of bias using ROBIS tool.** *J Clin Epidemiol* 2017 Sep 9. pii: S0895-4356(17)30534-6. doi: 10.1016/j.jclinepi.2017.08.015. [Epub ahead of print]



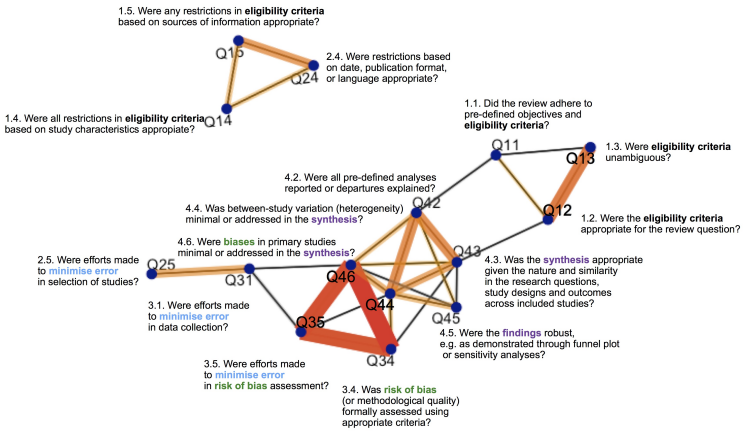
Suppl. Fig. 1: PRISMA flow diagram of article selection process.



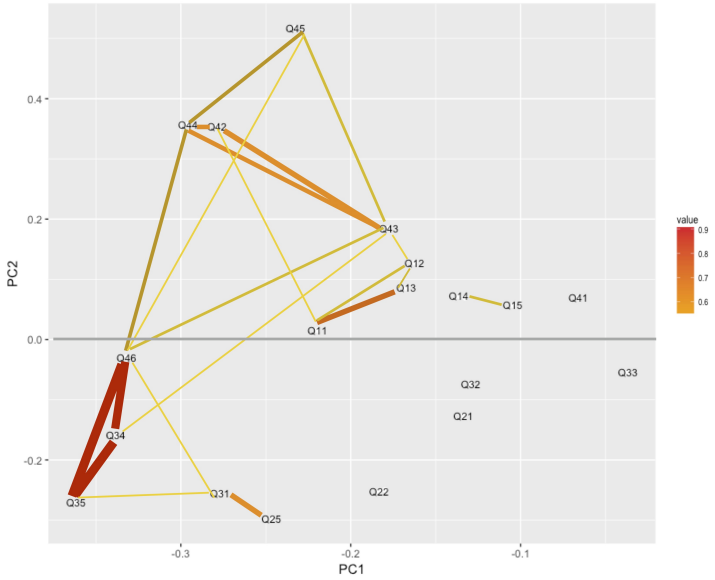
Suppl. Fig. 2: Percentage of explained variance by dimension and variable or individual contributions on PC1/PC2 for results of AMSTAR- and ROBIS-based principal component analyses (PCAs). Fig. S2a and S2b show top-10 PCA dimensions sorted by percentage of explained variance in AMSTAR (S2a) and ROBIS (S2b) based PCAs. Fig. S2c and S2d represent variable contributions on PC1/PC2 of AMSTAR (S2c) and ROBIS (S2d) based PCAs. Fig. S2e and S2f show top 30 individual contributions on PC1/PC2 of AMSTAR (S2c) and ROBIS (S2d) based PCAs.



Suppl. Fig. 3: Correlation matrix of pair responses comparation to signaling questions of phase 2 ROBIS tool.



Suppl. Fig. 4: Network plot of highly correlated phase 2 ROBIS signaling questions. Edges connect nodes (signaling questions) that are highly correlated. The color (from yellow to red) and the width (from >0.5 to 1) of the edges represent the grade of correlation between connected nodes.



Suppl. Fig. 5: Network plot of highly correlated phase 2 ROBIS signaling questions mapped using ROBIS-PCA derived PC1/PC2 coordinates. Edges connect nodes (signaling questions) that are highly correlated. Nodes are represented in a coordinated system defined by PC1 and PC2 components of ROBIS-based PCA. Those nodes mapped far from (0,0) coordinates contribute more to the observed variability. The color (from yellow to red) and the width (from >0.5 to 1) of the edges represent the grade of correlation between connected nodes.